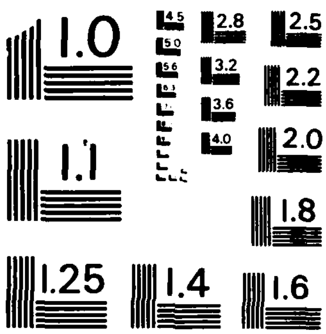
AD-A169 717 OUTLIER RESISTANT FILTERING AND SMOOTHING(U) 1/1
CONNECTICUT UNIV STORRS DEPT OF ELECTRICAL ENGINEERING
AND CO.. H TSAKNAKIS ET AL. APR 86 UCT/DEECS/TR-86-5
UNCLASSIFIED AFOSR-TR-86-0365 AFOSR-83-0229 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

DTIC ELECTE JUL 2 4 1986 S D D

(2)

REPORT DOCUMENTATION PAGE

AD-A169 717

1b. RESTRICTIVE MARKINGS	
2b.	NA
3. DISTRIBUTION/AVAILABILITY OF REPORT	Approved for Public Release; Distribution Unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	UCT/DEECS/TR-86-5
5. MONITORING ORGANIZATION REPORT NUMBER(S)	AFOSR-TR- 86-0365
6a. NAME OF PERFORMING ORGANIZATION	University of Connecticut
6b. OFFICE SYMBOL (If applicable)	
7a. NAME OF MONITORING ORGANIZATION	AFOSR/NM
6c. ADDRESS (City, State and ZIP Code)	U-157, 260 Glenbrook Rd. Storrs, CT 06268
7b. ADDRESS (City, State and ZIP Code)	Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	AFOSR
8b. OFFICE SYMBOL (If applicable)	NM
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	AFOSR-83-0229
8c. ADDRESS (City, State and ZIP Code)	Bldg. 410 Bolling AFB, DC

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304	A5	

11. TITLE (Include Security Classification) Outlier Resistant Filtering & Smoothing

12. PERSONAL AUTHOR(S) H. Tsaknakis and P. Papantoni-Kazakos

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
interim	FROM 7/1/85 TO 6/30/86	April 1986	~~56~~ 67

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
XXX	XXXXXXXXX	XXXXXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Filtering, Smoothing, Qualitative robustness, Outlier Resistance.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We consider a stationary Gaussian information process transmitted through an additive noise channel. ~~We~~ assume that the noise and information processes are mutually independent, and ~~we~~ model the noise process as nominally Gaussian with additive independent outliers. For the above system model, ~~we~~ first develop a theory for outlier resistant filtering and smoothing operations. ~~We~~ then design specific such nonlinear operations, and ~~we~~ study their performance. The performance criteria are asymptotic mean squared error at the Gaussian model, the breakdown point, and the influence function. ~~We find~~ that our operations combine excellent at the nominal model performance, with strong resistance to outliers.

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Prof. P. Papantoni-Kazakos	(203) 486-2867	AFOSR/NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 86-0365

OUTLIER RESISTANT FILTERING AND SMOOTHING

by
Haralampos Tsaknakis and P. Papantoni-Kazakos
University of Connecticut
EECS Department, U-157
Storrs, Connecticut 06268

Abstract

We consider a stationary Gaussian information process transmitted through an additive noise channel. We assume that the noise and information processes are mutually independent, and we model the noise process as nominally Gaussian with additive independent ouliers. For the above system model, we first develop a theory for outlier resistant filtering and smoothing operations. We then design specific such nonlinear operations, and we study their performance. The performance criteria are the asymptotic mean squared error at the Gaussian nominal model, the breakdown point, and the influence function. We find that our operations combine excellent at the nominal model performance with strong resistance to outliers.

This work was supported by the U.S. Air Force Office of Scientific Research, under the grant AFOSR-83-0229.

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

1. Introduction

In filtering and smoothing, information carrying data are extracted from noisy observations. The formalization and solution of the filtering and smoothing problems are well established, when the joint process that characterizes the relationship between information and noise data sequences is statistically well known (see Kalman (1960, 1963), Kolmogorov (1941), and Wiener (1949)), or parametrically known. Linear filtering and smoothing operations are then by far the most widely used, due to their simplicity in implementation. In practice, however, the occurence of occasional extremely erroneous data values, called outliers, are frequently observed. Furthermore, linear data operations are notoriously nonresistant to such outliers, inducing dramatic performance instabilities. The purpose of this paper is to establish a theory for outlier resistant filtering and smoothing procedures and to provide specific such data operations for Gaussian information processes, and additive, nominally Gaussian, noise processes. Our presentation is based on the theory of qualitative robustness (see Boente et al (1982), Cox (1978), Hampel (1971), Papantoni-Kazakos and Gray (1979), and Papantoni-Kazakos (1981, 1983, 1984a, 1984b)).

2. Preliminaries

Let us consider discrete-time information and noise stochastic processes. Let $X_i^k = \{X_i,..,X_k\}$, $W_i^k = \{W_i,..,W_k\}$, and $Y_i^k = \{Y_i,...,Y_k\}$, $i \geq k$, denote random data sequences respectively generated by the information, noise, and observation (joint information and noise) stochastic processes. Let $x_i^k = \{x_i,..,x_k\}$, $w_i^k = \{w_i,..,w_k\}$, and $y_i^k = \{y_i,..,y_k\}$, $i \geq k$, denote realizations of respectively the random sequences, X_i^k, W_i^k, and Y_i^k, and let x_i^k, w_i^k, and y_i^k all take their values on the Euclidean space R^{k-i+1}. Let μ denote the measure of the observation process, and let μ^n denote the n-dimensional restriction of μ. Let the objective be to estimate the information datum x_ℓ, from the observation sequence y_i^k, where $i \leq \ell \leq k$, and let the mean squared criterion be used. If the measure, μ, is well known, then the mean squared estimate,

$\hat{x}_\ell$, of x_ℓ is given by the conditional expectation $E\{X_\ell | y_i^k, \mu^{k-i+1}\}$. If the measure μ is also Gaussian, then the latter conditional expectation is a linear transformation of the observation sequence y_i^k. Given the measure, μ, and the observation sequence, y_i^k, the mean squared estimate, $\hat{x}_\ell = \hat{x}_\ell(y_i^k)$, is a function of the sequence y_i^k, whose specific form is determined by the measure μ. The induced mean squared error, $e_{ik}(\mu,\hat{x}_\ell)$, is then equal to the expected value $E\{[X_\ell - \hat{x}_\ell(Y_i^k)]^2 | \mu^{k-i+1}\}$, where $\hat{x}_\ell(y_i^k) = E\{X_\ell | y_i^k, \mu^{k-i+1}\}$.

The occurence of occasional outliers induces uncertainties in the description of the measure μ. The initial issue is then the qualitative characterization of those uncertainties, with the final objective being the design of outlier resistant filtering and smoothing operations. As it has been previously established (Boente et al (1982) and Papantoni-Kazakos (1983, 1984b)), the outlier model is best described by a Prohorov class. In particular, if μ_o denotes the nominal measure of the observation sequences, the outlier model is described by the Prohorov ball $\Pi_{n,\rho_n}(\mu_o,\mu) \le \varepsilon$ of processes, where ε represents frequency of outlier occurence, and where $\rho_n(x^n,y^n)$ is a distortion measure between data sequences x^n and y^n respectively generated by the measures μ_o and μ, defined as follows:

$$\rho_n(x^n,y^n) = \begin{cases} n^{-1} \sum_i |x_i - y_i| \triangleq \gamma_n(x^n,y^n) \; ; \text{ for given finite } n \\ \\ \inf\{\alpha : n^{-1}[\# i : \gamma_m(x_{i+1}^{i+m}, y_{i+1}^{i+m}) > \alpha] \le \alpha\}; \text{ for} \end{cases} \tag{1}$$

$n > n_o$, where m and n_o are fixed positive integers

If N denotes the class of joint processes, ν, whose marginals are μ_o and μ, then the Prohorov distance $\Pi_{n,\rho_n}(\mu_o,\mu)$ is defined as follows:

$$\Pi_{n,\rho_n}(\mu_o,\mu) = \inf_{\nu \in N} \; \inf\{\delta : \nu(x^n, y^n : \rho_n(x^n, y^n) > \delta) \le \delta\} \tag{2}$$

The distortion measure $\rho_n(x^n, y^n)$ in (1) is clearly a metric, for all n, the integer m is a design parameter corresponding to outlier patterns, and n_o is an integer determined by the nominal measure μ_o, for the satisfaction of performance

stability. In particular, given n, an estimate, $\hat{x}_\ell\,(y_i^{i+n-1})$, $i\leq\ell$, of the information datum x_ℓ is called <u>outlier resistant</u> or <u>qualitatively robust</u> at μ_o, iff:

Given $\varepsilon>0$, there exists $\delta>0$, such that,

$$\Pi_{n,\rho_n}(\mu_o,\mu)<\delta \text{ implies } |e_{i,i+n-1}(\mu_o,\hat{x}_\ell)-e_{i,i+n-1}(\mu,\hat{x}_\ell)|<\varepsilon$$

Furthermore, for the estimate $\hat{x}_\ell\,(y_i^{i+n-1})$, $i\leq\ell$ to be outlier resistant at μ_o, for given finite n, it is sufficient that $\hat{x}_\ell$ be bounded, and that the following continuity condition holds (see Papantoni-Kazakos and Gray (1979), Boente et al (1982), and Papanto Kazakos (1981, 1983, 1984a, 1984b)), if the class N in (2) includes stationary processes only.

<u>Condition A</u>

For given finite n, pointwise continuity as a function of the data. That is, given $\varepsilon>0$, given x_i^{i+n-1}, there exists $\delta>0$, such that,

$$y_i^{i+n-1}:\ n^{-1}\sum_{j=i}^{i+n-1}|x_j-y_j|<\delta \text{ implies } |\hat{x}_\ell(x_i^{i+n-1})-\hat{x}_\ell(y_i^{i+n-1})|<\varepsilon$$

If N is a class of stationary and ergodic processes, then the limit $\lim_{n\to\infty}\Pi_{n,\rho_n}(\mu_o,\mu)$ of the Prohorov distance in (2) equals the Prohorov distance $\Pi_{m,\gamma_m}(\mu_o,\mu)$, where $\gamma_m(x^m,y^m)$ is as in (1). In addition, if the measure μ_o^m is absolutely continuous, and if then $f_o^m(x^m)$ denotes the density function induced by the measure μ_o^m at the vector point x^m, the class F^m of density functions defined below is contained in the class $\Pi_{m,\gamma_m}(\mu_o,\mu)\leq\varepsilon$ of measures μ.

$$F^m = \{\ f^m = (1-\varepsilon)\ f_o^m + \varepsilon h^m\ ,\ h^m \text{ any m-dimensional density function}\} \qquad (3)$$

The class F^m of densities in (3) represents the occurence of arbitrary outlier m-dimensional data sequences, with probability ε, where with probability $1-\varepsilon$ each m-dimensional data sequence is generated by the nominal measure μ_o.

3. The Model and Outline of the Approach

We will adopt the nominal model of additive and mutually independent information and noise processes, which possess density functions, for any given dimensionalit n. Given n, these density functions are respectively denoted f^n_{os} and f^n_{oN}, and the corresponding density function, f^n_o, of the nominal observation process is then given by the convolution $f^n_{os} * f^n_{oN}$. We will consider m-dimensional outlier patterns, as described by the class F^m in (3), with $f^m_o = f^m_{os} * f^m_{oN}$. Given the above model, we wish to estimate the data sequences, $\ldots, x_{-1}, x_o, x_1, \ldots$, generated by the density function f_{os}. In particular, each information datum x_ℓ is estimated via an observation sequence $y_i^{i+k-1}, i \leq \ell \leq i+k-1$, where if $\ell = i+k-1$ the estimation operation, $\hat{x}_\ell(y_i^{i+k-1})$, corresponds to filtering, and where it corresponds to smoothing otherwise.

Initially adopting the mean squared criterion, we are seeking filtering and smoothing operations that are outlier resistant, and simultaneously induce satisfactory performance at the nominal model.For finite number of observations, condition A,section in conjuction with boundness are sufficient for outlier resistance. They define, however, a large class of possible operations, and they do not incorporate performance at the nominal model. We will address the overall issue (outlier resistance and performance at the nominal model) via a combination of saddle point game theory and the theory of qualitative robustness. We will do this in two steps: (a) when each information datum is estimated via a length -k observation sequence, for $k \leq m$, where m as in (3), and (b) when the latter length k exceeds m. At each step, a saddle point game is first formalized and solved. Then, the induced by the latter solution filtering or smoothing operation is checked against the continuity condition A in section 2 and against boundness. If those conditions are not satisfied, the operations are appropriately modified. To avoid vagueness in our presentation, we assume from now on that the nominal information process (with density function f_{os}) is zero mean and Gaussian.

4. Construction of Filtering and Smoothing Operations-Step 1

Consider the model in section 3, with f_{os} Gaussian, zero mean, and not necessaril stationary. Let then $R_s(i,j)$ denote the autocovariance matrix of the random information data sequence X_i^j, where $j\geq i$, and consider the mean squared estimation of the the information datum x_ℓ, using an observation sequence y_i^{i+k-1}, such that $i\leq\ell\leq i+k-1$ and $k\leq m$, where m is as in (3). For dimensionalities $k\leq m$, the observation sequence y_i^{i+k-1} is generated by the k-dimensional restriction of some density function f^m in the class F^m. Given some estimate, $\hat{x}_\ell(y_i^{i+k-1})$, of the datum x_ℓ, the mean squared error, $e_{i,i+k-1}(f^m,\hat{x}_\ell)$, that it induces at the density f^m, equals $E\{[X_\ell-\hat{x}_\ell(Y_i^{i+k-1})]^2|f^m\}$. Given $k\leq m$, we then consider a saddle point game, with payoff function the error $e_{i,i+k-1}(f^m,\hat{x}_\ell)$. In particular, we search for a pair $(f_*^m,\hat{x}_\ell^*)$ such that $f_*^m \in F^m$, and,

$$\forall f^m \in F^m;\; e_{i,i+k-1}(f^m,\hat{x}_\ell^*) \leq e_{i,i+k-1}(f_*^m,\hat{x}_\ell^*) \leq e_{i,i+k-1}(f_*^m,\hat{x}_\ell);\; \forall \hat{x}_\ell \tag{4}$$

From the results in Papantoni-Kazakos (1984a), we conclude that for the class F^m in (3), the game in (4) has a saddle point $(f_*^m,\hat{x}_\ell^*)$. From the theory of saddle point games we then conclude that the saddle point can be found as follows.

$$(f_*^m,\hat{x}_\ell^*) \; : \; e_{i,i+k-1}(f_*^m,\hat{x}_\ell^*) = \sup_{f^m\in F^m} \inf_{\hat{x}_\ell} e_{i,i+k-1}(f^m,\hat{x}_\ell) \tag{5}$$

;where, if f^k denotes the k-dimensional restriction of the density f^m, then,

$$\inf_{\hat{x}_\ell} e_{i,i+k-1}(f^m,\hat{x}_\ell) = \sigma_\ell^2 - I(f^k) \tag{6}$$

$$\sigma_\ell^2 = E\{X_\ell^2\} \tag{7}$$

$$I(f^k) \triangleq E\{ E^2\{X_\ell | Y_i^{i+k-1}, f^k\} | f^k\} \tag{8}$$

;and thus,

$$f_*^m \; : \; I(f_*^k) = \inf_{f^m\in F^m} I(f^k) \tag{9}$$

$$\hat{x}_\ell^* \; : \; \hat{x}_\ell^*(y_i^{i+k-1}) = E\{ X_\ell | y_i^{i+k-1}, f_*^k\} \tag{10}$$

Let us now denote by V_ℓ the $(\ell-i+1)$th row of the autocovariance matrix $R_s(i,i+k-1)$; that is, $V_\ell = [E\{X_iX_\ell\},\ldots,E\{X_{i+k-1}X_\ell\}]$. Let now $\nabla_\ell f^k(y_i^{i+k-1})$ denote the directional derivative of $f^k(y_i^{i+k-1})$ with respect to the column vector V_ℓ^T. Then, since the information process is zero mean Gaussian, we easily find that the quantity $I(f^k)$ in (8) takes the following form, where R denotes

the real line.

$$I(f^k) = \int_{R^k} \frac{[\nabla_\ell f^k(y_i^{i+k-1})]^2}{f^k(y_i^{i+k-1})} \, d\, y_i^{i+k-1} \qquad (11)$$

The functional in (11) is a generalized Fisher information measure, and it is vaguely lower semicontinuous in the vague topology of all the k-dimensional density functions (Huber(1981)). In addition, the closure, F_c^m, of the class F^m in (3) is vaguely compact. Therefore, the infimum, $\inf_{f^m F^m} I(f^k)$, exists and is attained in F_c^m. If f_*^m is some density in F^m that attains the latter infimum, then the estimate x_ℓ^* in (10) takes the following form.

$$\hat{x}_\ell^* : \hat{x}_\ell^* (y_i^{i+k-1}) = - \frac{\nabla_\ell f_*^k(y_i^{i+k-1})}{f_*^k(y_i^{i+k-1})} \qquad (12)$$

Regarding uniqueness, we proceed with the following theorem, whose proof is in appendix A.

<u>Theorem 1</u>

Let f_1^k and f_2^k be two densities in F^m that both attain the infimum $\inf_{f^m \varepsilon F^m} I(f^k)$. Then,

$$\frac{\nabla_\ell f_1^k(y_i^{i+k-1})}{f_1^k(y_i^{i+k-1})} = \frac{\nabla_\ell f_2^k(y_i^{i+k-1})}{f_2^k(y_i^{i+k-1})} \quad ; \text{a.e. } y_i^{i+k-1} \,\varepsilon R^k$$

The saddle point estimate $\hat{x}_\ell^*$ in (12) is thus a.e. unique <u>in F^m.</u>

To find the explicit form of a saddle point pair $(f_*, \hat{x}_\ell^*)$, we will now assume that the nominal noise process is zero mean Gaussian and we will denote the auto-covariance matrix of the random sequence $W_i^j, j \geq i$ from this process, $R_N(i,j)$. Considering the $(\ell-i+1)$ th row V_ℓ of the autocovariance matrix $R_s(i,i+k-1)$, defined earlier, we then denote,

$$M_{ik} = R_s\,(i,i+k-1) + R_N\,(i,i+k-1)$$

$$P_{ik}(\ell) = M_{ik}^{-1}\,V_\ell^T \tag{13}$$

$$r_{ik}(\ell) = V_\ell\,P_{ik}\,(\ell)$$

$$y^k = y_i^{i+k-1}$$

Denoting by $\phi(x)$ and $\Phi(x)$ respectively the density function and the cumulative distribution at the point x, of the zero mean, unit variance Gaussian random variable, we then express the following lemma, whose proof is in appendix A.

<u>Lemma 1</u>

Let the nominal information and noise processes in class F^m be both Gaussian, with zero mean and autocovariance matrices as above. Then, a saddle point solution $(f_*^k,\hat{x}_\ell^*)$ of the game in (4) is given by the expressions below, where $\hat{x}_\ell^*(y^k)$ is unique a.e. in R^k, and where $|\ |$, T, (-1) respectively denote determinant, transpose, and inverse.

$$f_*^k\,(y^k) = \begin{cases} (1-\varepsilon)\ (2\pi)^{-k/2}\ |M_{ik}|^{-1/2}\ \exp\left\{-\frac{1}{2}\,y^{k^T} M_{ik}^{-1}\,y^k\right\};\ y^k : |P_{ik}^T(\ell)y^k| \le \lambda \\ (1-\varepsilon)\ (2\pi)^{-k/2}\ |M_{ik}|^{-1/2}\ \exp\left\{\frac{\lambda^2}{2r_{ik}(\ell)} - \frac{1}{2}\left[y^{k^T} M_{ik}^{-1}\,y^k - \frac{\left(P_{ik}^T(\ell)y^k\right)^2}{r_{ik}(\ell)}\right] - \lambda\,\frac{|P_{ik}^T(\ell)y^k|}{r_{ik}(\ell)}\right\};\ y^k : |P_{ik}^T(\ell)y^k| > \lambda \end{cases} \tag{14}$$

$$\hat{x}_\ell^*\,(y^k) = \begin{cases} P_{ik}^T\,(\ell)\ y^k\ ;\ y^k : |P_{ik}^T\,(\ell)\ y^k| \le \lambda \\ \lambda\ \mathrm{sgn}\left(P_{ik}^T\,(\ell)\ y^k\right)\ ;\ y^k : |P_{ik}^T\,(\ell)\ y^k| > \lambda \end{cases} \tag{15}$$

; where,

$$\text{sgn}x = \begin{cases} 1 \ ; \ x \geq 0 \\ -1 \ ; \ x < 0 \end{cases}$$

$$\lambda = c\sqrt{r_{ik}(\ell)} \tag{16}$$

$$c : \Phi(c) + c^{-1}\phi(c) = \frac{2-\varepsilon}{2(1-\varepsilon)} \tag{17}$$

We note that given ε, the constant c in (17) is unique. Also, $r_{ik}(\ell)=I(f_o^k)$, where f_o^k is the nominal Gaussian observation density. Thus, given ℓ and i, $r_{ik}(\ell)$ is monotonically nondecreasing with increasing k. Similarly, given ℓ and k, $r_{ik}(\ell)$ is monotonically nondecreasing with decreasing i. We observe that the estimate in (15) is a truncated version of the linear, optimal at the nominal Gaussian model, mean squared estimate. In addition, for $\varepsilon \to 0$, the constant c in (17), tends to infinity, and so does then the truncation constant λ. In the latter case, the estimate in (15) becomes identical to the optimal at the nominal Gaussian model mean squared estimate. Finally, for any $\varepsilon>0$ in (17), the estimate in (15) is clearly bounded, and satisfies condition A in section 2; it is thus outlier resistent.

The mean squared error, $e_{i,i+k-1}(f_*^m,\hat{x}_\ell^*)$, induced by the estimate $\hat{x}_\ell^*$ in (15), at the least favorable in F^m density function $f_*^k(y^k)$ in (14), equals $\sigma_\ell^2-I(f_*^k)$, where $I(f^k)$ is the information measure in (11). The above error is the largest mean squared error induced by $\hat{x}_\ell^*$ in F^m, and by substitution we obtain,

$$e_{i,i+k-1}\left(f_*^m,\ \hat{x}_\ell^*\right) = \sigma_\ell^2\left\{1-(1-\varepsilon)\left(2\Phi(c)-1\right)\rho_{ik}^2(\ell)\right\} \tag{18}$$

; where c is as in (17), and $r_{ik}(\ell)$ is as in (13), and where,

$$\rho_{ik}(\ell) \triangleq \sigma_\ell^{-1}\sqrt{r_{ik}(\ell)} \tag{19}$$

Let $e_{i,i+k-1}\left(f_0^m,\hat{x}_\ell^*\right)$ denote the mean squared error induced by the estimate $\hat{x}_\ell^*$ in (15), at the nominal Gaussian observation density f_0^m, and let $e^0_{i,i+k-1}(\ell)$ denote the mean squared error induced by the optimal at f_0^m mean squared estimate of x_ℓ, at f_0^m, given the observation vector y_i^{i+k-1}. Then, via straight forward computations, and for c as in (17) and $\rho_{ik}(\ell)$ as in (19), we obtain,

$$e^0_{i,i+k-1}(\ell) = \sigma_\ell^2 \left[1-\rho_{ik}^2(\ell)\right] \tag{20}$$

$$e_{i,i+k-1}(f_0^m,\hat{x}_\ell^*) = e^0_{i,i+k-1}(\ell) + 2r_{ik}(\ell)\left(\Phi(-c)(1+c^2) - c\phi(c)\right) \tag{21}$$

5. Construction of Filtering and Smoothing Operations - Step 2

For the same model as in section 4, we now consider the case where the length of the observation sequence is larger than the integer m in (3). Considering observation sequences y_1^n ; we will distinguish here between causal filtering and noncausal filtering or smoothing. In the former, the information datum x_k is estimated, given the observation sequence y_1^k . In the latter, the datum x_k is estimated, when the sequence y_1^{2k-1} is observed.

5.1. Causal Filtering

Let us define,

$$\begin{aligned} V_k &= \left[E\{X_1X_k\},\dots, E\{X_kX_k\}\right] \\ M_k &= R_s(1,k) + R_N(1,k) \\ P_k &= M_k^{-1} V_k^T \\ r_k &= V_k P_k \end{aligned} \tag{22}$$

Then, for $k \leq m$, and directly from lemma 1, we conclude that the estimate $\hat{x}_k (y_1^k)$ is as follows.

$$\hat{x}_k \triangleq \hat{x}_k \left(y_1^k\right) = g_{F,k} \left(P_k^T y_1^k\right) \; ; \; k \leq m \tag{23}$$

; where,

$$g_{F,k} (z) = \begin{cases} z & ; \; |z| \leq \lambda_k \\ \lambda_k \operatorname{sgn} z & ; \; |z| > \lambda_k \end{cases} \tag{24}$$

$$\lambda_k = c \; r_k^{1/2} \tag{25}$$

$$c : \Phi (c) + c^{-1} \; \phi(c) = \frac{2-\varepsilon}{2(1-\varepsilon)} \tag{26}$$

Let us now assume that $k>m$, and let us consider the estimation of the datum x_k, given the estimates $\left\{\hat{x}_\ell (y_1^\ell); 1 \leq \ell \leq k-1\right\}$, and given the observation sequence y_1^k. For $\hat{x}_1^{k-\ell} : (\hat{x}_1^{k-\ell})^T = [\hat{x}_1, .., \hat{x}_{k-\ell}]$, $k-\ell \geq 1$, let us define,

$$\begin{aligned} \hat{x}_{I\ell} &= G^T_{\ell,\ell+m-k} R^{-1}_{s,k-m} \hat{x}_1^{k-m} \quad ; \; 1 \leq \ell \leq k-1 \\ \hat{x}_{Ik} &= G^T_{km} \; R^{-1}_{s,k-m} \; \hat{x}_1^{k-m} \end{aligned} \tag{27}$$

; where,

$$G^T_{km} = [E\{X_1 X_k\}, \ldots, E\{X_{k-m} X_k\}] \; , \; G^T_\ell \triangleq G^T_{\ell 1} \tag{28}$$

$$R_{s,k-\ell} = R_s(1,k-\ell) \tag{29}$$

Given $\ell : 1 \leq \ell \leq k-1$, the effective observation datum corresponding to the time index ℓ is then, $y_\ell - \hat{x}_{I\ell}$. In addition, given the sequence of estimates, $\left\{\hat{x}_\ell (y_1^\ell); 1 \leq \ell \leq k-m\right\}$, an initial estimate of the datum x_k is provided by $\hat{x}_{Ik}$. For the final estimate of the datum x_k ; $k > m$, we use the operation $g_{F,m}(z)$ in (24), in conjunction with the above effective observations, to obtain,

$$\hat{x}_k \triangleq \hat{x}_k(y_1^k) = G_{km}^T \; R_{s,k-m}^{-1} \; \hat{x}_1^{k-m} + g_{F,m}([y_k - \hat{x}_{Ik},\ldots,y_{k-m+1} - \hat{x}_{I,k-m+1},0,\ldots,0]P_k) \quad ; \; k > m \qquad (30)$$

; where $G_k^T, R_{s,k-1}$, and P_k are respectively as in (28), (29), and (22), and where,

$$\left(\hat{x}_1^{k-\ell}\right)^T = \left[\hat{x}_1,\ldots,\hat{x}_{k-\ell}\right] , \; \hat{x}_\ell \triangleq \hat{x}_\ell\left(y_1^\ell\right) \qquad (31)$$

As an alternative, we also consider the following causal filtering operation.

$$\hat{x}_{kA} = \begin{cases} g_{F,k}\left(P_k^T \; y_1^k\right); \; k \leq m \\ B_{mk}^T \; R_s^{-1}(m,k-1) \; \hat{z}_m^{k-1} \quad ; \; k > m \end{cases} \qquad (32)$$

;where, $\left(\hat{z}_m^{k-1}\right)^T = \left[\hat{z}_m,\ldots,\hat{z}_{k-1}\right]$, $B_{mk}^T = [E\{X_m X_k\},\ldots,E\{X_{k-1} X_k\}]$, and,

$$\hat{z}_\ell = g_{F,m}\left(P_m^T \; y_{\ell-m+1}^\ell\right) \; ; \; \ell \geq m \qquad (33)$$

5.2 Noncausal Filtering or Smoothing

Let us define,

$$V_{k\ell} = \left[E\{X_{k-\ell}X_k\} \; ,\ldots, \; E\{X_{k+\ell} \; X_k\}\right]$$

$$M_{k\ell} = R_s\;(k-\ell,\; k+\ell) + R_n(k-\ell,k+\ell)$$

$$P_{k\ell} = M_{k\ell}^{-1} \; V_{k\ell}^T \qquad (34)$$

$$r_{k\ell} = V_{k\ell} \; P_{k\ell}$$

For $2\ell+1 \leq m$, and directly from lemma 1, we then conclude,

$$\hat{x}_{k\ell} \triangleq \hat{x}_k\left(y_{k-\ell}^{k+\ell}\right) = g_{S,k,\ell}\left(P_{k\ell}^T \; y_{k-\ell}^{k+\ell}\right) \quad ; \; 2\ell \leq m-1 \qquad (35)$$

; where c is as in (26), and,

$$g_{S,k,\ell}(z) = \begin{cases} z & ; \ |z| \le \mu_{k\ell} \\ \mu_{k\ell}\ \mathrm{sgn}(z) & ; \ |z| > \mu_{k\ell} \end{cases} \tag{36}$$

$$\mu_{k\ell} = c\ r_{k\ell}^{1/2} \tag{37}$$

For ℓ such that $2\ell > m - 1$, and $m = 2n + 1$, we consider the following smoothing operation.

$$\hat{x}_{kA} = \begin{cases} g_{S,k,k-1}\left(P^T_{k,k-1} y_1^{2k-1}\right) & ; \ 2(k-1) \le m-1 = 2n \\ H^T_{k,k-n-1}\ \Lambda^{-1}_{s,k,k-n-1}\ \hat{W}_{k-m} & ; \ 2(k-1) > m-1 = 2n \end{cases} \tag{38}$$

; where,

$$(\hat{W}_{k-m})^T = [\ (\hat{w}^{k-1}_{n+1})^T\ ,\ (\hat{w}^{2k-n-1}_{k+1})^T\ ]$$

$$\left(\hat{w}^{k-1}_{n+1}\right)^T = \left[\hat{w}_{n+1,n}, \ldots, \hat{w}_{k-1,n}\right] \tag{39}$$

$$\left(\hat{w}^{2k-n-1}_{k+1}\right)^T = \left[\hat{w}_{k+1,n}, \ldots, \hat{w}_{2k-n-1,n}\right]$$

$$\hat{w}_{\ell,n} = g_{S,\ell,n}\left(P^T_{\ell,n}\ y^{\ell+n}_{\ell-n}\right)$$

6. Asymptotic Properties

In this section, we are focusing on the asymptotic properties of the operations presented in section 5. We begin by proving that the latter operations are outlier resistant, for asymptotically long data sequences. For finite length such sequences, the operations are outlier resistant, since they satisfy condition A in section 2. Regarding asymptotic outlier resistance,

we will go directly to the definition of outlier resistance, stated in section 2. We will consider sequences of i.i.d. m-size blocks of outlier values, that are additive to and independent of the nominal Gaussian noise process. We then express the following theorems, whose proofs are in appendix B.

Theorem 2

Let the Gaussian information process in sections 4 and 5 be such that: There exists some finite positive constant c, such that,

$$
\begin{aligned}
& E\{X_k^2\} \le c \quad ; \forall k \\
& |H_{k,k-1}^T \Lambda_{s,k,k-1}^{-1} I_{2(k-1)}| \le c \; ; \forall k \qquad (40) \\
& |G_k^T R_{s,k-1}^{-1} I_{k-1}| \le c \; ; \forall k
\end{aligned}
$$

;where $I_n^T \triangleq [1,\ldots,1]$, with n elements, and where G_k^T and $R_{s,k-1}$ are respectively given by (28) and (29).

Then, the operation in (30) is asymptotically ($k \to \infty$) outlier resistant, within the class of stationary and ergodic observation processes, and for mutually independent m-size batches of ouliers.

Theorem 3

Let the Gaussian information process satisfy conditions (40) in theorem 2. Then, for mutually independent m-size batches of outliers, the operations in (32) and (39) are outlier resistant, for every k.

We will now turn to the evaluation of the asymptotic at the nominal Gaussian model mean squared error, induced by the operations in (32),(38), and (30), in the case where the Gaussian nominal information and noise processes are both stationary, with respective power spectral densities, $f_s(\omega)$ and $f_v(\omega)$, $\omega\epsilon[-\pi,\pi]$.

We will start with the operations in (32) and (38), whose asymptotic forms, when the information datum x_0 is estimated from infinite past and future observation data, are then respectively as follows.

$$\hat{x}_{0F} = \sum_{\ell=-\infty}^{-1} d_\ell \, g_{F,m} \left(P_m^T \, y_{\ell-m+1}^{\ell}\right) \tag{41}$$

$$\hat{x}_{0S} = \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} a_\ell \, g_{S,\ell,n}\left(P_{\ell,n}^T y_{\ell-n}^{\ell+n}\right) \tag{42}$$

The sets $\{d_\ell\}$ and $\{a_\ell\}$ are respectively the prediction and interpolation coefficients corresponding to the Gaussian nominal information process. The functions $g_{F,m}(\cdot)$ and $g_{S,\ell,n}(\cdot)$ are respectively as in (24) and (36), and they both induce stationary processes, when operating on sequences from the stationary nominal Gaussian model. Let us then denote by $f_{g_{F,m}}(\omega)$ and $f_{g_{S,n}}(\omega)$, $\omega\epsilon$ $[-\pi,\pi]$, the power spectral densities induced respectively by $g_{F,m}(\cdot)$ and $g_{S,\ell,n}(\cdot)$ and the stationary Gaussian nominal model, where then $g_{S,\ell,n}(\cdot)$ is not a function of ℓ. Let $D(\omega)$ and $A(\omega)$, $\omega\epsilon$ $[-\pi,\pi]$, be the Fourier transforms of respectively the sequences $\{d_\ell;-\infty<\ell<-1\}$ and $\{a_\ell;-\infty<\ell<\infty,\ell\neq 0\}$, and let us define,

$$H_{F,m}(\omega) = [1,e^{-j\omega},\ldots,e^{-j\omega(m-1)}]P_m \tag{43}$$

$$H_{S,n}(\omega) = [e^{j\omega n},\ldots e^{j\omega},1,e^{-j\omega},\ldots e^{-j\omega n}]P_n \tag{44}$$

; where P_m in (43) is as in (22), and where P_n in (44) is as P_{kn} in (34), for any k in the case of the stationary Gaussian nominal model. Given all the above, and denoting by $e(f_o, \hat{x}_{0F})$ and $e(f_o, \hat{x}_{0S})$ the mean squared errors induced respectively by the estimates in (41) and (42), at the nominal stationary Gaussian observation density f_o, we obtain,

$$e(f_o, \hat{x}_{0F}) = (2\pi)^{-1}\left\{ \int_{-\pi}^{\pi} f_s(\omega)d\omega - 2[2\Phi(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}(D(\omega)H_{F,m}(\omega))f_s(\omega)d\omega \right.$$

$$\left. + \int_{-\pi}^{\pi} f_{g_{F,m}}(\omega)\,||D(\omega)||^2 d\omega \right\} \tag{45}$$

$$e(f_o, \hat{x}_{0S}) = (2\pi)^{-1}\left\{ \int_{-\pi}^{\pi} f_s(\omega)d\omega - 2[2\Phi(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}(A(\omega)H_{S,n}(\omega))f_s(\omega)d\omega \right.$$

$$\left. + \int_{-\pi}^{\pi} f_{g_{S,n}}(\omega)\,||A(\omega)||^2 d\omega \right\} \tag{46}$$

; where Re(.) denotes real part, where $||h(\omega)||^2 = h(\omega)h^*(\omega)$, where $\Phi(x)$ is the Gaussian distribution at the point x, and where the constant c is as in (26).

The quantities that present problems in the evaluation of the asymptotic mean squared errors, in (45) and (46), are the power spectral densities $f_{g_{F,m}}(\omega)$ and $f_{g_{S,n}}(\omega)$. These power spectral densities correspond respectively to the sequences, $\{g_{F,m}(Z_i)\}$ and $\{g_{S,n}(E_i)\}$ of random variables, where $Z_i = P_m^T Y_{i-m+1}^i$ and $E_i = P_n^T Y_{i-n}^{i+n}$, and where for c as in (26), λ_m as in (25), and $\mu_n = \mu_{kn}$; $\forall k$ as in (37), the functions $g_{F,m}(z)$ and $g_{S,n}(z)$ are as follows.

$$g_{F,m}(z) = \begin{cases} z; |z| \le \lambda_m \\ \lambda_m \mathrm{sgn}(z); |z| > \lambda_m \end{cases} \tag{47}$$

$$g_{S,n}(z) = \begin{cases} z; |z| \le \mu_n \\ \mu_n \mathrm{sgn}(z); |z| > \mu_n \end{cases} \tag{48}$$

We will seek upper and lower bounds on the errors in (45) and (46), via the derivation of such bounds on the spectral characteristics of the sequences $\{g_{F,m}(Z_i)\}$ and $\{g_{S,n}(E_i)\}$. We note that the power spectral densities, $f_Z(\omega)$

and $f_E(\omega)$, of respectively the Gaussian sequences $\{Z_i = P_m^T Y_{i-m+1}^i\}$ and $\{E_i = P_n^T Y_{i-n}^{i+n}\}$, when $\{Y_i\}$ is generated by the Gaussian stationary nominal density f_o, are easily found to be given by expressions (49) and (50) below, where $H_{F,m}(\omega)$ and $H_{S,n}(\omega)$ are respectively as in (43) and (44).

$$f_Z(\omega) = ||H_{F,m}(\omega)||^2 [f_s(\omega) + f_N(\omega)] ; \omega \varepsilon [-\pi, \pi] \tag{49}$$

$$f_E(\omega) = ||H_{S,n}(\omega)||^2 [f_s(\omega) + f_N(\omega)] ; \omega \varepsilon [-\pi, \pi] \tag{50}$$

Let us now define the following quantities.

$$\begin{aligned} & r_{pF} \triangleq E\{Z_i Z_{i+p} | f_o\} \quad , \quad r_{pS} \triangleq E\{E_i E_{i+p} | f_o\} \\ & \sigma_F^2 \triangleq r_{0F} \quad , \quad \gamma_{pF} \triangleq \frac{r_{pF}}{\sigma_F^2} \quad , \quad \sigma_{pS}^2 \triangleq \sigma_F^2 \, [1-\gamma_{pF}^2] \\ & \sigma_S^2 \triangleq r_{0S} \quad , \quad \gamma_{pS} \triangleq \frac{r_{pS}}{\sigma_S^2} \quad , \quad \sigma_{pS}^2 \triangleq \sigma_S^2 \, [1-\gamma_{pS}^2] \end{aligned} \tag{51}$$

$$\begin{aligned} b_{pA}(n) \triangleq\; & r_{pA} \left[2\, \Phi \left(\frac{c}{[1-\gamma_{pA}^2]^{1/2}} \right) - 1 \right] \left[2\, \Phi\, (c) - 1 \right] \\ & + 2U\,(n-2) r_{pA} \sum_{k=1}^{n-1} \left[\gamma_{pA}^{-2} - 1 \right]^{-k} \phi^{(2k-1)} \left(\frac{c}{[1-\gamma_{pA}^2]^{1/2}} \right) \cdot \\ & \cdot \left\{ \frac{2\Phi(c)-1}{2^k \, k!} + \sum_{\ell=1}^{k} c^{2\ell-1} \, 2^{k-\ell+1} \left[\frac{k!}{(2k+1)!\,(\ell-1)!} - 2^{-2(k-\ell+1)} \frac{(\ell-1)!}{k!\,(2\ell-1)!} \right] \right\} ; \\ & ; A = F \text{ or } S \end{aligned} \tag{52}$$

$$U(x) = \begin{cases} 1 \; ; & x > 0 \\ 0 \; ; & x < 0 \end{cases}$$

$$B_{N,\{n_p\},A}(\omega) \triangleq [2\Phi(c)-1]^2 f_s(\omega) + \sum_{1\le|p|\le N-1} e^{j\omega p}\left\{b_{pA}(n_p) - [2\Phi(c)-1]^2 r_{pA}\right\}$$

$$+ \sigma_A^2 \{2\Phi(c)[3-c^2-2\Phi(c)]-2c\phi(c) + 2c^2-2\} \text{ ; } A=F \text{ or } S \tag{53}$$

$$e^{\ell}_{N,\{n_p\},F}(f_o,\hat{x}_0) \triangleq (2\pi)^{-1}\left\{ \int_{-\pi}^{\pi} f_s(\omega)d\omega - 2[2\Phi(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}(D(\omega)H_{F,m}(\omega))f_s(\omega)d\omega + \int_{-\pi}^{\pi} ||D(\omega)||^2 B_{N,\{n_p\},F}(\omega)d\omega\right\} \tag{54}$$

$$e^{\ell}_{N,\{n_p\},S}(f_o,\hat{x}_0) \triangleq (2\pi)^{-1}\left\{ \int_{-\pi}^{\pi} f_s(\omega)d\omega - 2[2\Phi(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}(A(\omega)H_{S,n}(\omega))f_s(\omega)d\omega + \int_{-\pi}^{\pi} ||A(\omega)||^2 B_{N,\{n_p\},S}(\omega)d\omega\right\} \tag{55}$$

; where $D(\omega)$ and $A(\omega)$ are respectively as in (45) and (46), where $H_{F,m}(\omega)$ and $H_{S,n}(\omega)$ are respectively as in (43) and (44), and where c is as in (26). Then, for $e(f_o,\hat{x}_{0F})$ and $e(f_o,\hat{x}_{0S})$ respectively as in (45) and (46), we can express the following theorem, whose proof is in appendix B.

Theorem 4

Let r_{pF} and r_{pS} in (51) be such that, $r_{|p|,F} > r_{|p|+1,F} \searrow 0$ and $r_{|p|,S} > r_{|p|+1,S} \searrow 0$. Let γ_{pF} and γ_{pS} in(59) be such that, $\gamma_{pF} < e(1+e)^{-1}$ and $\gamma_{pS} < e(1+e)^{-1}$, $\forall\ |p| \ge 1$. Then, given $\delta > 0$, there exist positive finite integers, N_F and N_S, and sets of positive integers, $\{n_{pF}\ ;\ 1 \le |p| \le N_F-1\}$ and $\{n_{pS}\ ;\ 1 \le |p| \le N_S-1\}$, such that, $n_{|p|,F} > n_{|p|+1,F}$ and $n_{|p|,S} > n_{|p|+1,S}$; $\forall$ p, and,

$$|e(f_o,\hat{x}_{0F}) - e^{\ell}_{N_F,\{n_{pF}\},F}(f_o,\hat{x}_0)|<\delta$$

$$|e(f_o,\hat{x}_{0S}) - e^{\ell}_{N_S,\{n_{pS}\},S}(f_o,\hat{x}_0)|<\delta$$

; where $e^{\ell}_{N_F,\{n_{pF}\},F}(f_o,\hat{x}_0)$ and $e_{N_S,\{n_{pS}\},S}(f_o,\hat{x}_0)$ are given respectively by (54) and

(55). The latter errors are then used as close approximations of respectively the mean squared errors $e(f_o, \hat{x}_{0F})$ and $e(f_o, \hat{x}_{0S})$, in (45) and (46).

Let us now turn to the estimate induced by the operation in (30). For stationary information and noise processes, and for the datum x_0 being estimated from infinite past and future observation data, the asymptotic form of this estimate is as follows.

$$\hat{x}_{0F} = \sum_{\ell=-\infty}^{-m} d_{\ell m} \hat{x}_{\ell F} + g_{F,m} \left(\sum_{i=-m+1}^{0} b_i [y_i - \sum_{j=-\infty}^{-m} d_{j,m+i} \hat{x}_{jF}] \right) \tag{56}$$

; where $\{d_{\ell p}\}$ are the prediction coefficients corresponding to the Gaussian nominal information process, for predicting x_0 from $\{x_\ell ; -\infty < \ell \leq -p\}$, and where $\{b_i\}$ are the filtering coefficients at the nominal Gaussian model, for estimating x_0 from $\{y_\ell ; -\infty < \ell \leq 0\}$. The function $g_{F,m}(\cdot)$ is given by (24), with k=m.

We will study the asymptotic performance of the prediction estimate in (56), for autoregressive Gaussian nominal information processes, and additive white Gaussian nominal noise processes. Let the nominal model be described as follows.

$$\begin{aligned} \underline{u}_n &= A\underline{u}_{n-1} + B\underline{v}_{n-1} \\ y_n &= B^T \underline{u}_n + w_n \end{aligned} \tag{57}$$

; where the sequences $\{V_i\}$ and $\{W_i\}$ are mutually independent, i.i.d. and zero mean Gaussian, with respective variances σ^2 and r^2, and where,

$$\underline{u}_n^T = [x_n, x_{n-1}, \ldots, x_{n-k+1}]$$

$$A = \begin{bmatrix} a_1, & a_2, & \ldots, & a_k \\ 1, & 0, & \ldots, & 0 \\ 0, & 1, & 0, \ldots, & 0 \\ \vdots & & & \\ 0, & \ldots\ldots, & 0, & 1 \end{bmatrix} \tag{58}$$

$$B^T = [1, 0, \ldots, 0]$$

Then, the estimate in (56) takes the following form.

$$\underline{\hat{u}}_{0F} = A^{m}\, \underline{\hat{u}}_{-m,F} + \underline{g}_{F,m}\left(\sum_{i=-m+1}^{0} \underline{b}_i \left[y_i - B^T A^{m+i}\, \hat{u}_{-m,F} \right] \right) \tag{59}$$

; where $\{\underline{b}_i\}$ are the vector coefficients of the optimal at the nominal model linear mean squared estimate of $\underline{u}_0$, given $\{y_i \; ; \; i \le 0\}$, and where for c as in (17), and for $r_{-j,m}$ being the variance gain induced by the optimal at the nominal model linear mean squared estimate of x_{-j}, given $\{y_i \; ; \; -m+1 \le i \le 0\}$, we have,

$$\begin{aligned}
&\underline{g}^T_{F,m}(\underline{z}) = \left[g_{F,0,m}(z_1), \ldots, g_{F,-k+1,m}(z_k)\right] \\
&\underline{z}^T = [z_1, \ldots, z_k] \\
&g_{F,-j,m}(x) = \begin{cases} x & ; \; |x| \le \lambda_{-j,m} \\ \lambda_{-j,m}\, \mathrm{sgn}\,(x) \; ; & |x| > \lambda_{-j,m} \end{cases} \\
&\lambda_{-j,m} = c\sqrt{r_{-j,m}}
\end{aligned} \tag{60}$$

Let us now define,

$$\begin{aligned}
&\underline{Z} \triangleq \sum_{i=-m+1}^{0} \underline{b}_i \left[Y_i - B^T A^{m+i}\, \underline{U}_{-m} \right] \\
&S_g \triangleq E\left\{ \underline{g}_{F,m}(\underline{Z})\; \underline{g}^T_{F,m}(\underline{Z}) \right\} \\
&S_U \triangleq E\left\{ [\underline{U}_0 - A^m\, \underline{U}_{-m}][\underline{U}_0 - A^m\, \underline{U}_{-m}]^T \right\} \\
&S_{uz} \triangleq E\left\{ \underline{Z}\, [\underline{U}_0 - A^m\, \underline{U}_{-m}]^T \right\} = \left\{ \tau_{ij};\; i,j=1,..k \right\} \\
&S_z \triangleq E\left\{ \underline{Z}\, Z^T \right\} = \left\{ s^2_{ji} \; ; \; i,\; j=1,\ldots,k \right\} \\
&S_{oF} \triangleq E\left\{ [U_0 - u_{0F}]\; [U_0 - \hat{u}_{0F}]^T \right\} \\
&P = S_u - FS_{uz} - S^T_{uz}F + S_g \; , \quad F = \{f_{ij} : f_{ii} = 2\Phi(s^{-1}_{ii}[\lambda_{-i+1,m}]) - 1 \; , \; f_{ij} = 0 \; ; \; i \ne j\} \\
&C \triangleq \sum_{i=-m+1}^{0} b_i\, B^T A^i
\end{aligned} \tag{61}$$

$$G_{ij} \triangleq E\{g_{F,1-i,m}(Z_i) g_{F,1-j,m}(Z_j) [\underline{Z}\,\underline{Z}^T - S_z]\}$$

$$G'_{ij} \triangleq E\{\underline{e}_i^T S_{uz}^T S_z^{-1} \underline{Z}\, g_{F,1-i,m}(Z_j)\ [Z\,Z^T - S_z]\}$$

$$N_{ij} \triangleq [2\,\Phi\left(\frac{\lambda_{-j+1,m}}{s_{jj}}\right) - 1]\ \underline{e}_j\ \underline{e}_i^T\ [S_{uz}^T\ S_z^{-1}\ C - I] +$$

$$+ [2\,\Phi\left(\frac{\lambda_{-i+1,m}}{s_{ii}}\right) - 1]\ \underline{e}_i\ \underline{e}_j^T\ [S_{uz}^T\ S_z^{-1}\ C - I]$$

$$Q_{ij} \triangleq 2^{-1}[C^T\ S_z^{-1}\left(G'_{ij} + G'_{ji} - G_{ij}\right) S_z^{-1}\ C - C^T N_{ij} - N_{ij}^T C]$$

; where $i,j = 1,\ldots,k$, where $\underline{e}_i \triangleq [0,..0,\overset{\overset{i}{\uparrow}}{1},0,..0]^T$, and where the expectations are taken at the nominal model. We then express the following theorem, whose proof is in appendix B.

Theorem 5

Let the process $\{\underline{U}_n\}$ be asymptotically stationary, and let S_{oF}^u and S_{oF}^{ℓ} be the solutions of the following matrix equations.

$$S_{oF}^u = A^m\ S_{oF}^u\ (A^T)^m + P$$
$$S_{oF}^{\ell} = A^m\ S_{oF}^{\ell}\ (A^T)^m + P - Q \tag{62}$$

; where, $Q = \{\ \mathrm{tr}(A^m\ S_{oF}^{\ell}\ [A^T]^m\ Q_{ij})\ ,\ i,j = 1,\ldots k\}$

Then,

(i) The solutions S_{oF}^{ℓ} and S_{oF}^u exist.

(ii) The residual process $\underline{U}_0 - \hat{\underline{U}}_0$ is asymptotically stationary.

If $M_1 \leq M_2$ means that each diagonal element of the matrix M_1 is bounded from above by the corresponding diagonal element of the matrix M_2, then,

$$S_{oF}^{\ell} \leq S_{oF} \leq S_{oF}^u \tag{63}$$

(iii) If tr denotes trace, and if $\mu_{max}(A)$ is the largest magnitude eigenvalue of the matrix A, then,

$$| \mu_{max} (A) | < 1$$

$$| \mathrm{tr} (S^{u}_{oF} - S^{\ell}_{oF}) | \leq \frac{|\mu_{max}(A)|^{2m}}{1 - |\mu_{max}(A)|^{2m}} (\mathrm{tr} P) \qquad (64)$$

If the nominal model in (57) is first order autoregressive, with autoregressive parameter $\alpha < 1$, then the matrices in (61) and (62) reduce to scalars, and the bounds in (63) are easily found to be as follows, where c is as in (17).

$$\begin{aligned} s^{\ell}_{oF} &= p \, [1 - \alpha^{2m} q]^{-1} \\ s^{u}_{oF} &= p \, [1 - \alpha^{2m}]^{-1} \end{aligned} \qquad (65)$$

; where, for $\sigma^2 \triangleq E\{V_0^2\}$ and $r^2 \triangleq E\{W_0^2\}$ in (57),

$$\begin{aligned}
\lambda_m &\triangleq \lambda_{0,m} = c\sqrt{r_{0,m}} \\
s_u^2 &\triangleq E\left\{[X_0 - \alpha^m X_{-m}]^2\right\} = (1-\alpha^2)^{-1}\sigma^2(1-\alpha^{2m}) \\
s_{11}^2 &\triangleq E\left\{[\sum_{i=-m+1}^{0} b_i (Y_i - \alpha^{m+i} X_{-m})]^2\right\} = r^2 \sum_{i=-m+1}^{0} b_i^2 + \frac{\sigma^2}{1-\alpha^2} \left\{ \sum_{i=-m+1}^{0} \sum_{j=-m+1}^{0} b_i b_j \alpha^{|i-j|} - \alpha^{2m} d^2 \right\} \\
s_{uz} &\triangleq E\left\{ [X_0 - \alpha^m X_{-m}] \sum_{i=-m+1}^{0} b_i (Y_i - \alpha^{m+i} X_{-m}) \right\} = \sum_{i=-m+1}^{0} b_i \alpha^{-i} - \alpha^{2m} d \\
f &\triangleq 2\,\Phi\left(\frac{\lambda_m}{s_{11}}\right) - 1 \\
d &\triangleq \sum_{i=-m+1}^{0} b_i \alpha^i \\
p &\triangleq \lambda_m^2 + s_u^2 + (s_{11}^2 - \lambda_m^2 - 2s_{uz}) f - 2\lambda_m s_{11} \phi\left(\frac{\lambda_m}{s_{11}}\right) \\
q &\triangleq 1 - d(2-d)f - 2d^2 s_{11}^{-1} \lambda_m [1 - s_{uz} s_{11}^{-1}] \phi\left(\frac{\lambda_m}{s_{11}}\right)
\end{aligned} \qquad (66)$$

The inequality in (64) takes then the following form:

$$| s^{u}_{oF} - s^{\ell}_{oF} | \leq \frac{\alpha^{2m}}{1 - \alpha^{2m}} \, p \tag{67}$$

The scalars p and tr P, respectively in (67) and (64), are both bounded for every value of the truncation parameters $\{\lambda_{j,m}\}$. The bounds in (64) and (67) are thus exponentially decreasing to zero, with increasing m. The lower and upper bounds in (63) and (65) are therefore approaching each other with exponential rate, as the design parameter m increases.

7. Performance Measures for Outlier Resistance

Let us consider the frequently observed in practice case of indeoendent and additive outliers. In particular, let the noise sequence $\{\ldots, W_{-1}, W_0, W_1, \ldots\}$ be such that each of its elements is generated by the nominal Gaussian noise process, with probability $1-\delta$, and it is instead equal to some deterministic value, v, with probability δ, $0 \leq \delta \leq 1$. Let the value v occur with probability δ, independently per noise datum. Given the above outlier model, given some asymptotic filtering or smoothing operation, $\hat{x}_0$, let $e(f_o,\delta,v,\hat{x}_0)$ denote the induced mean squared error. That is, if f_o represents the overall nominal Gaussian model, then, $e(f_o,\delta,v,\hat{x}_0) = E\{[X_0-\hat{x}_0]^2 | f_o,\delta,v\}$. Let us denote, $e(f_o,\delta,\hat{x}_0) \triangleq \lim_{v \to +\infty} e(f_o,\delta,v,\hat{x}_0)$, and let there exist some value δ^*, $0 \leq \delta^* \leq 1$, such that,

$$e(f_o,\delta,\hat{x}_0) > E\{X_0^2 | f_o\} \;\; ; \; \forall \delta > \delta^*$$

$$e(f_o,\delta,\hat{x}_0) \leq E\{X_0^2 | f_o\} \;\; ; \; \forall \delta \leq \delta^*$$

Then, the value δ^* is called the <u>breakdown point</u> of the asymptotic operation $\hat{x}_0$. The breakdown point clearly represents the maximum frequency of independent, asymptotically large in amplitude outliers that the operation $\hat{x}_0$ can tolerate, before it becomes worthless; that is, before it starts inducing mean squared error, that is larger than that deduced when no observation data are available.

We note that the breakdown points of the optimal at the Gaussian models linear filtering and smoothing operations, are easily found to equal zero.

Let us now consider a generalization of the outlier model presented above. In particular, let us consider the case where independent, size m blocks of outliers may occur. Then, each block occurs with probability δ, and it consists of a value v per datum in the block. Given some filtering or smoothing operation $\hat{x}_0$, we then denote the induced mean squared error, $e_m(f_o,\delta,v,\hat{x}_0)$. Denoting by $e(f_o,\hat{x}_0)$ the mean squared error in the absence of the above outlier model, we denote, $J_{m,\delta}(v) \stackrel{\Delta}{=} e_m(f_o,\delta,v,\hat{x}_0) - e(f_o,\hat{x}_0)$. We call $J_{m,\delta}(v)$ the variation function at δ. Given δ, the variation function exhibits the difference between the mean squared error, when the outlier value is v and the frequency of the outlier blocks is δ, and the mean squared error in the absence of outliers. We call $I_{m,\delta}(v) \stackrel{\Delta}{=} \delta^{-1} J_{m,\delta}(v)$, the normalized variation function at δ, and we call $I_m(v) \stackrel{\Delta}{=} \lim_{\delta\to 0} I_{m,\delta}(v)$ the influence function. The influence function is the slop of the variation function at $\delta = 0$, and it exhibits the effect of the outlier value v, at asymptotically small outlier frequencies δ. It is easily found that the optimal at the Gaussian model linear filtering and smoothing operations induce, for m=1, influence function, $I_1^o(v)$, that is given by the following expression, where $C(\omega)$ is the Fourier transform of the linear filter or smoother, and where $f_N(\omega)$ denotes the power spectral density of the Gaussian noise.

$$I_1^o(v) = (2\pi)^{-1} \left\{ \left[v^2 + (2\pi)^{-1} \int_{-\pi}^{\pi} f_N(\omega)d\omega\right] \int_{-\pi}^{\pi} \|C(\omega)\|^2 d\omega - 2\int_{-\pi}^{\pi} \|C(\omega)\|^2 f_N(\omega)d\omega \right\} \tag{68}$$

As a function of the outlier value v, the influence function $I_1^o(v)$ is quadratic, and it increases to infinity as $v \to +\infty$.

In this section, we study the breakdown points and the influence functions of the operations in (32) and (38), and the operation in (30). We adopt respectively the same nominal Gaussian models as in section 6, and we start with the operations in (32) and (38). Let us then define the spectral densities $f_s(\omega)$, $f_{g_{S,n}}(\omega)$, and $f_{g_{F,m}}(\omega)$, and the Fourier transforms $A(\omega)$ and $D(\omega)$ as in section 6. Let $H_{F,m}(\omega)$ and $H_{S,n}(\omega)$ be as in (43) and (44), and let us define,

$$r_{g_{F,m}}(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{g_{F,m}}(\omega)\, e^{-j\omega k} d\omega$$

$$r_{g_{S,n}}(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{g_{S,n}}(\omega)\, e^{-j\omega k} d\omega \tag{69}$$

$$U(x) = \begin{cases} 1 \ ; & x \geq 0 \\ 0 \ ; & x < 0 \end{cases}$$

Then, after some straightforward transformations, we find the following mean squared expressions, for c as in (17).

$$\begin{aligned} e(f_o,\delta,\hat{x}_{0F}) = {} & (2\pi)^{-1} \int_{-\pi}^{\pi} f_s(\omega)d\omega + \lambda_m^2 D^2(0) - \\ & - (1-\delta)^m \Big\{ 2[2\Phi(c)-1]\ (2\pi)^{-1} \int_{-\pi}^{\pi} \mathrm{Re}(D(\omega)H_{F,m}(\omega)) f_s(\omega)d\omega \\ & + 2\lambda_m^2 D^2(0) - [\lambda_m^2 + r_{g_{F,m}}(0)]\ (2\pi)^{-1} \int_{-\pi}^{\pi} ||D(\omega)||^2 d\omega \Big\} \\ & + U(m-2) \sum_{k=1}^{m-1} (1-\delta)^{m+k} [r_{g_{F,m}}(k) + \lambda_m^2](2\pi)^{-1} \int_{-\pi}^{\pi} ||D(\omega)||^2 [e^{-j\omega k} + e^{j\omega k}]d\omega \\ & + (1-\delta)^{2m}(2\pi)^{-1} \Big\{ \lambda_m^2 D^2(0) - \lambda_m^2 \int_{-\pi}^{\pi} ||D(\omega)||^2 [\sum_{k=-m+1}^{m-1} e^{j\omega k}]d\omega + \\ & + \int_{-\pi}^{\pi} ||D(\omega)||^2 f_{g_{F,m}}(\omega)d\omega - \int_{-\pi}^{\pi} ||D(\omega)||^2 [\sum_{k=-m+1}^{m-1} r_{g_{F,m}}(k) e^{j\omega k}]d\omega \Big\} \end{aligned} \tag{70}$$

$$
\begin{aligned}
e(f_o,\delta,\hat{x}_{0S}) = {} & (2\pi)^{-1}\int_{-\pi}^{\pi} f_s(\omega)d\omega + \mu_n^2 A^2(0) - \\
& - (1-\delta)^{2n+1}\Big\{ 2[2\Phi(c)-1](2\pi)^{-1}\int_{-\pi}^{\pi} R_e(A(\omega)H_{S,n}(\omega))f_s(\omega)d\omega \\
& + 2\mu_n^2 A^2(0) - [\mu_n^2 + r_{g_{S,n}}(0)](2\pi)^{-1}\int_{-\pi}^{\pi} ||A(\omega)||^2 d\omega \Big\} \\
& + U(n-1)\sum_{k=1}^{2n}(1-\delta)^{2n+1+k}[r_{g_{S,n}}(k) + \mu_n^2](2\pi)^{-1}\int_{-\pi}^{\pi} ||A(\omega)||^2[e^{-j\omega k} + e^{j\omega k}]d\omega \\
& + (1-\delta)^{2(2n+1)}(2\pi)^{-1}\Big\{ \mu_n^2 A^2(0) - \mu_n^2\int_{-\pi}^{\pi} ||A(\omega)||^2[\sum_{k=-2n}^{2n} e^{j\omega k}]d\omega + \\
& + \int_{-\pi}^{\pi} ||A(\omega)||^2 f_{g_{S,n}}(\omega)d\omega - \int_{-\pi}^{\pi} ||A(\omega)||^2[\sum_{k=-2n}^{2n} r_{g_{S,n}}(k)\, e^{j\omega k}]d\omega \Big\}
\end{aligned} \tag{71}
$$

; where λ_m and μ_n are respectively as in (47) and (48). Let us now consider the quantities $\{b_{pA}(n)\}$, in (52), and $B_{N,\{n_p\},A}(\omega)$, in (53), and let us denote,

$$
H_A(\omega) = \begin{cases} H_{F,m}(\omega) & ; \text{ for } A=F \\ H_{S,n}(\omega) & ; \text{ for } A=S \end{cases}
$$

$$
C_A(\omega) = \begin{cases} A(\omega) & ; \text{ for } A=S \\ D(\omega) & ; \text{ for } A=F \end{cases} \tag{72}
$$

$$
\nu_A^2 = \begin{cases} \lambda_m^2 & ; \text{ for } A=F \\ \mu_n^2 & ; \text{ for } A=S \end{cases}
$$

$$
n_A = \begin{cases} m & ; \text{ for } A=F \\ 2n+1 & ; \text{ for } A=S \end{cases}
$$

Let us then define,

$$F_A(\delta) \triangleq \nu_A^2\, c_A^2(0) - (1-\delta)^{n_A}\Big\{ 2[2\Phi(c)-1](2\pi)^{-1}\int_{-\pi}^{\pi} \mathrm{Re}(C_A(\omega)H_A(\omega))f_s(\omega)d\omega \; +$$

$$+\; 2\nu_A^2 c_A^2(0) - [\nu_A^2 + b_{0A}(n_0)](2\pi)^{-1}\int_{-\pi}^{\pi} ||C_A(\omega)||^2 d\omega \;\Big\}$$

$$+\; U(n_A-2)\sum_{k=1}^{n_A-1}(1-\delta)^{n_A+k}[\nu_A^2 + b_{kA}(n_k)](2\pi)^{-1}\int_{-\pi}^{\pi} ||C_A(\omega)||^2[e^{j\omega k} + e^{-j\omega k}]d\omega$$

$$+\; (1-\delta)^{2n_A}(2\pi)^{-1}\Big\{ \nu_A^2 c_A^2(0) - \nu_A^2\int_{-\pi}^{\pi} ||C_A(\omega)||^2[\sum_{k=-n_A+1}^{n_A-1} e^{j\omega k}]d\omega \; +$$

$$+\int_{-\pi}^{\pi} ||C_A(\omega)||^2 B_{N,\{n_p\},A}(\omega)d\omega - \int_{-\pi}^{\pi} ||C_A(\omega)||^2[\sum_{k=-n_A+1}^{n_A-1} b_{kA}(n_k)e^{j\omega k}]d\omega \Big\}$$

$$; \; A = F \text{ or } S \qquad (73)$$

We now express the following theorem, whose proof is in appendix C.

Theorem 6

Let the conditions in Theorem 4 be satisfied. Let $\delta^*_{F,m}$ and $\delta^*_{S,n}$ denote then the breakdown points of respectively the filtering and smoothing operations in (32) and (38). Then, given $\zeta > 0$, there exist positive finite integers N_F and N_S, and sets of positive integers, $\{n_{pF} \; ; \; 1 \le |p| \le N_F-1\}$ and $\{n_{pS} \; ; \; 1 \le |p| \le N_S-1\}$, such that, $n_{|p|,F} > n_{|p|+1,F}$ and $n_{|p|,S} > n_{|p|+1,S}$; $\forall$ p, and such that when substituted in respectively $F_F(\delta)$ and $F_S(\delta)$, they give,

(i) Unique roots of the functions $F_F(\delta)$ and $F_S(\delta)$, respectively denoted $\delta^o_{F,m}$ and $\delta^o_{S,n}$.

(ii) $|\delta^*_{F,m} - \delta^o_{F,m}| < \zeta$

$|\delta^*_{S,n} - \delta^o_{S,n}| < \zeta$

The breakdown points $\delta^*_{F,m}$ and $\delta^*_{S,n}$ are unique, and they are both strictly larger than zero.

We point out that the filtering and smoothing operations in (32) and (38) induce bounded, for every value v, variation and influence functions. We do not include the computation of bounds on the latter functions, due to the tediousness in their derivation. We now proceed with the study of the breakdown point and the influence function of the filtering operation in (30), for the nominal model in (57). For the latter model, we first compute the influence function, $I^O_m(v)$, induced by the optimal at the nominal model filtering operation. Then, we study the breakdown point and the influence function, $I_m(v)$, induced by the outlier resistant filtering operation in (30). Let C be the kxk matrix defined in (61), and let us define,

$$\underline{\mu} \triangleq \sum_{i=-m+1}^{0} \underline{b}_i = \{\mu_i\}$$

$$N \triangleq \sum_{i=-m+1}^{0} \underline{b}_i \, \underline{b}_i^T \qquad (74)$$

Then, we can express the following lemma, where for the nominal model in (57), $r^2 \triangleq E\{W_0^2\}$. The proof of the lemma is in appendix C.

<u>Lemma 2</u>

Let the nominal model in (57) be asymptotically stationary. Then, the influence function $I^O_m(v)$ is given asymptotically by the following expression, where I denotes the kxk identity matrix.

$$I^O_m(v) \triangleq \sum_{i=0}^{\infty} [\,(I-C)\,A^m]^i\,[v^2\,\underline{\mu}\,\underline{\mu}^T - r^2N]\,[\,(A^T)^m\,(I-C)^T]^i \qquad (75)$$

We note that for the scalar form of the model in (57), with autoregressive parameter $\alpha < 1$, the influence function in (75) reduces to the following expression.

$$I^O_m(v) \triangleq \frac{v^2\,\mu^2 - r^2\,\nu}{1 - \alpha^{2m}\,(1-\xi)^2} \qquad (76)$$

; where,

$$\nu \triangleq \sum_{i=-m+1}^{0} b_i^2$$

$$\xi \triangleq \sum_{i=-m+1}^{0} b_i \, \alpha^i \qquad (77)$$

$$\mu \triangleq \sum_{i=-m+1}^{0} b_i$$

Let us consider the quantities defined in (61) and the vector $\underline{\mu}$ defined in (74), and let us in addition define,

$$\underline{Z}_1 \triangleq \sum_{i=-m+1}^{0} \underline{b}_i \, B^T[\underline{U}_i - A^{m+i}\underline{U}_{-m}] = [Z_{11},\ldots,Z_{1k}]^T$$

$$S_{z_1} \triangleq E\{\underline{Z}_1 \, \underline{Z}_1^T\} = \{\rho_{ij}^2 \; ; \; i,j=1,\ldots,k\}$$

$$S_{g,v} \triangleq E\{\underline{g}_{F,m}(\underline{Z}_1 + v\underline{\mu}) \, \underline{g}_{F,m}^T(\underline{Z}_1 + v\underline{\mu})\}$$

$$f_{i,v} \triangleq \Phi\left(\frac{v\mu_i + \lambda_{-i+1,m}}{\rho_{ii}}\right) - \Phi\left(\frac{v\mu_i - \lambda_{-i+1,m}}{\rho_{ii}}\right)$$

$$F_v \triangleq \{f_{ij,v} : f_{ii,v} = f_{i,v} \, , \, f_{ij,v} = 0 \; ; \; i \neq j\} \qquad (78)$$

$$P_v \triangleq S_u - F_v S_{uz} - S_{uz}^T F_v + S_{g,v}$$

$$G_{ij,v} \triangleq E\{g_{F,1-i,m}(Z_{1i}+v\mu_i) \, g_{F,1-j,m}(Z_{1j}+v\mu_j)[\underline{Z}_1 \, \underline{Z}_1^T - S_{z_1}]\}$$

$$H_{ij,v} \triangleq E\{\underline{e}_i^T S_{uz}^T S_{z_1}^{-1} \underline{Z}_1 \, g_{F,1-j,m}(Z_{1j}+v\mu_j)[\underline{Z}_1 \, \underline{Z}_1^T - S_{z_1}]\}$$

$$N_{ij,v} \triangleq [f_{j,v} \, \underline{e}_j \, \underline{e}_i^T + f_{i,v} \, \underline{e}_i \, \underline{e}_j^T] \, [S_{uz}^T \, S_{z_1}^{-1} C - I]$$

$$Q_{ij,v} \triangleq 2^{-1}[C^T S_{z_1}^{-1} (H_{ij,v} + H_{ji,v} - G_{ij,v}) S_{z_1}^{-1} C - C^T N_{ij,v} - N_{ij,v}^T C]$$

$$\underline{\lambda}_m \triangleq [\lambda_{0,m},\ldots,\lambda_{-k+1,m}]^T$$

We can then express the following theorem, whose proof is in appendix C.

Theorem 7.

Let the nominal model in (57) be asymptotically stationary, and such that the elements of the matrix A are all nonnegative. Then,

(i) Let $I^{\ell}_{m}(v)$ and $I^{u}_{m}(v)$ be the respective solutions of the following matrix equations.

$$I^{\ell}_{m}(v) = P_v + A^m I^{\ell}_{m}(v)\ (A^T)^m - Q_v + A^m S^{\ell}_{oF}(A^T)^m - S^u_{oF}$$
$$I^{u}_{m}(v) = P_v + A^m I^{u}_{m}(v)\ (A^T)^m - S^{\ell}_{oF} + A^m S^{u}_{oF}(A^T)^m \tag{79}$$

; where, $Q_v \triangleq \{tr(A^m I^{\ell}_{m}(v)\ [A^T]^m\ Q_{ij}) + tr(A^m S^u_{oF}\ [A^T]^m\ Q_{ij,v})\ ; i,j=1,..,k\}$, and where S^{ℓ}_{oF} and S^{u}_{oF} are as in theorem 5.

Then, the influence matrix function, $I_m(v)$, induced by the filtering operation in (30) and the nominal model in (57), is bounded from above and below as follows.

$$tr\ I^{\ell}_{m}(v) \leq tr\ I_m(v) \leq tr\ I^{u}_{m}(v) \tag{80}$$

(ii) Let $S^u_m(\delta)$ and $S^{\ell}_m(\delta)$ be the respective solutions of the following matrix equations, given δ: $0 \leq \delta \leq 1$.

$$S^u_m(\delta) = A^m S^u_m(\delta)(A^T)^m + (1-\delta)^m P + [1-(1-\delta)^m][S_u + \underline{\lambda}_m \underline{\lambda}_m^T]$$
$$S^{\ell}_m(\delta) = A^m S^{\ell}_m(\delta)\ (A^T)^m - (1-\delta)^m [R-P] + [1-(1-\delta)^m][S_u + \underline{\lambda}_m \underline{\lambda}_m^T] \tag{81}$$

; where $R \triangleq \{tr(A^m S^{\ell}_m\ [A^T]^m Q_{ij})\ ;\ i,j=1,..,k\}$. Define, $S_o \triangleq E\{\underline{U}_0 \underline{U}_0^T | f_o\}$. Then, the functions $f_1(\delta) \triangleq tr(S^u_m(\delta) - S_o)$ and $f_2(\delta) \triangleq tr(S^{\ell}_m(\delta) - S_o)$ have unique roots, respectively denoted δ^u_m and δ^{ℓ}_m. Those roots are both strictly positive and less than one. Furthermore, the breakdown point, δ^*_m, of the operation in (30) is unique, and such that,

$$\delta^u_m \leq \delta^*_m \leq \delta^{\ell}_m \tag{82}$$

In the scalar case, where the nominal model in (57) is first order autoregressive, with autoregressive parameter α: $0 < \alpha < 1$, the results in theorem 7 simplify as follows.

$$\frac{p_v(1-q\alpha^{2m}) + p\alpha^{2m}}{(1-q\alpha^{2m})^2} - \frac{p(q_v + \alpha^{2m})}{(1-\alpha^{2m})(1-q\alpha^{2m})} \le I_m(v) \le \frac{p_v(1-\alpha^{2m}) + p\alpha^{2m}}{(1-\alpha^{2m})^2} - \frac{p}{(1-\alpha^{2m})(1-q\alpha^{2m})} \tag{83}$$

$$\delta_m^u \le \delta_m^* \le \delta_m^\ell \quad : \quad \begin{cases} \delta_m^u \text{ solution of : } f^u(\delta) \triangleq \dfrac{s_u^2 + \lambda_m^2 + (1-\delta)^m[p - s_u^2 - \lambda_m^2]}{1-\alpha^{2m}} - \dfrac{\sigma^2}{1-\alpha^2} = 0 \\ \delta_m^\ell \text{ solution of : } f^\ell(\delta) \triangleq \dfrac{s_u^2 + \lambda_m^2 + (1-\delta)^m[p - s_u^2 - \lambda_m^2]}{1-\alpha^{2m} + (1-\delta)^m\alpha^{2m}(1-q)} - \dfrac{\sigma^2}{1-\alpha^2} = 0 \end{cases} \tag{84}$$

; where $\sigma^2 \triangleq E\{V_0^2\}$ for V_0 as in (57), where $p, q, s_u, s_{uz}, \lambda_m$, and c are as in (66), where μ is as in (77), and where, for s_{11} as in (66) and $r^2 \triangleq E\{W_0^2\}$ in (57),

$$\rho^2 = E\{[\sum_{i=-m+1}^{0} b_i \; (X_i - \alpha^{m+i} X_{-m})]^2 \,|\, f_o\} = s_{11}^2 - r^2 \sum_{i=-m+1}^{0} b_i^2$$

$$f_v = \Phi(\frac{v\mu+\lambda_m}{\rho}) - \Phi(\frac{v\mu-\lambda_m}{\rho}) \tag{85}$$

$$p_v = s_u^2 + f_v[\rho^2 + v^2\mu^2 - \lambda_m^2 - 2s_{uz}] + 2\rho\mu v[\phi(\frac{v\mu+\lambda_m}{\rho}) - \phi(\frac{v\mu-\lambda_m}{\rho})]$$
$$+ \rho(v\mu-\lambda_m)\;\phi(\frac{v\mu-\lambda_m}{\rho}) - \rho(v\mu+\lambda_m)\;\phi(\frac{v\mu+\lambda_m}{\rho}) + \lambda_m^2$$

$$q_v = c(2-c)\; f_v + c^2\Big\{\frac{\lambda_m}{\rho}\,[1 - \frac{uz}{\rho^2}]\;[\phi(\frac{v\mu+\lambda_m}{\rho}) + (\frac{v\mu-\lambda_m}{\rho})]$$
$$- \frac{s_{uz}}{\rho^2}\,\mu v[\phi(\frac{v\mu+\lambda_m}{\rho}) - \phi(\frac{v\mu-\lambda_m}{\rho})]\Big\}$$

We conclude this section by presenting the form of the influence functions induced by the filtering and smoothing operations in (32) and (38), when the latter are modified to operate on disjoint rather than sliding block data. Let then $H_{F,m}(\omega)$ and $H_{S,n}(\omega)$ be as in (43) and (44), let c be as in (17), let λ_m and μ_n be as in (47) and (48), let $A(\omega)$ and $D(\omega)$ be as in section 6, let $f_s(\omega)$ be the power spectral density of the stationary information process, and let us define,

$$t_F^2 \triangleq E\,\{[P_m^T X_{i-m+1}^i]^2 \mid f_o\} \qquad t_S^2 \triangleq E\,\{[P_n^T X_{i-n}^{i+n}]^2 \mid f_o\}$$

$$p_{F,m}(v) \triangleq E\,\{[g_{F,m}(P_m^T X_{i-m+1}^i + vH_{F,m}(0)) - g_{F,m}(P_m^T Y_{i-m+1}^i)\,]^2 \mid f_o\}$$

$$p_{S,n}(v) \triangleq E\,\{[g_{S,n}(P_n^T X_{i-n}^{i+n} + v\,H_{S,n}(0)) - g_{S,n}(P_n^T Y_{i-n}^{i+n})]^2 \mid f_o\}$$

$$q_{F,p}(v) \triangleq E\,\{g_{F,m}(P_m^T Y_{i+m(p-1)+1}^{i+mp})\; g_{F,m}(P_m^T X_{i-m+1}^i + v\,H_{F,m}(0)) \mid f_o\} \tag{86}$$

$$q_{S,p}(v) \triangleq E\,\{g_{S,n}(P_n^T Y_{i+n(2p-1)+p}^{i+n(2p+1)+p})\; g_{S,n}\,(P_n^T X_{i+n}^{i+n} + v\,H_{S,n}(0)) \mid f_o\}$$

; where f_o represents the nominal model. Let the nominal noise process be stationary, let then $h_{g_{F,m}}(\omega,v)$ and $h_{g_{S,n}}(\omega,v)$ be the Fourier transforms of the sequences $\{q_{F,p}(v)\}$ and $\{q_{S,p}(v)\}$, and for Z_i and E_i as in (51) let $f_{g_{F,m}}(\omega)$ and $f_{g_{S,n}}(\omega)$ be respectively the Fourier transforms of the sequences

$\{E\{g_{F,m}(Z_i)\; g_{F,m}\,(Z_{i+mp}) \mid f_o\}\}$ and $\{E\{g_{S,n}(E_i) g_{S,n}(E_{i+(2n+1)p})\}\}$. Then, the influence functions of the corresponding filtering and smoothing operations are respectively given by the following expressions.

$$I_{F,m}(v) = \pi^{-1} \int_{-\pi}^{\pi} ||D(\omega)||^2 [h_{g_{F,m}}(\omega,v) - f_{g_{F,m}}(\omega) + 2^{-1}\, p_{F,m}(v)]\, d\omega$$
$$-\pi^{-1}[\,\Phi\,(\frac{\lambda_m - v\,H_{F,m}(0)}{t_F}) + \Phi\,(\frac{\lambda_m + v\,H_{F,m}(0)}{t_F}) - 2\,\Phi\,(c)]\cdot \tag{87}$$
$$\cdot \int_{-\pi}^{\pi} \mathrm{Re}(D(\omega)\;H_{F,m}(\omega))\; f_s(\omega) d\omega$$

$$I_{S,n}(v) = \pi^{-1} \int_{-\pi}^{\pi} ||A(\omega)||^2 [h_{g_{S,n}}(\omega,v) - f_{g_{S,n}}(\omega) + 2^{-1}\, p_{S,n}(v)] d\omega$$
$$-\,\pi^{-1}[\,\Phi(\frac{\mu_n - vH_{S,n}(0)}{t_S}) + \Phi\,(\frac{\mu_n + vH_{S,n}(0)}{t_S}) - 2\,\Phi\,(c)]\cdot$$
$$\cdot \int_{-\pi}^{\pi} \mathrm{Re}(A(\omega)\;H_{S,n}(\omega))\; f_s(\omega) d\omega \tag{88}$$

In contrast to the influence functions induced by the optimal at the nominal model linear filtering and smoothing operations (see (68)), the functions in (87) and (88) remain bounded for every v value. In fact, from the latter expressions we easily conclude,

$$\lim_{v \to \pm\infty} I_{F,m}(v) = \pi^{-1}[2\,\Phi(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}(D(\omega)H_{F,m}(\omega))\; f_s(\omega)d\omega +$$
$$+ \lambda_m^2\; [2^{-1} + \Phi\,(-c) - c^{-1}\,\phi(c)]\;\pi^{-1}\int_{-\pi}^{\pi} ||D(\omega)||^2 d\omega$$
$$- \pi^{-1}\int_{-\pi}^{\pi} ||D(\omega)||^2\; f_{g_{F,m}}(\omega)d\omega \tag{89}$$

$$\lim_{v \to \pm\infty} I_{S,n}(v) = \pi^{-1}[2\,\Phi\,(c)-1]\int_{-\pi}^{\pi} \mathrm{Re}\;(A(\omega)\;H_{S,n}(\omega))f_s(\omega)d\omega$$
$$+ \mu_n^2\;[2^{-1} + \Phi\,(-c)\;-c^{-1}\,\phi\,(c)]\;\pi^{-1}\int_{-\pi}^{\pi} ||A(\omega)||^2 d\omega$$
$$- \pi^{-1}\int_{-\pi}^{\pi} ||A(\omega)||^2\; f_{g_{S,n}}(\omega)d\omega \tag{90}$$

8. Numerical Results

In this section, we present some numerical results, regarding the performance of the filtering and smoothing operations in section 5. We consider the nominal model in (57), and we first evaluate the performance of the causal filtering operation in (59). Defining,

Σ : The solution of the matrix equation,

$$\Sigma = A[\Sigma-\Sigma BB^T\Sigma(B^T\Sigma B+r^2)^{-1}]A^T + \sigma^2\;BB^T$$

$$\underline{\beta} \triangleq \Sigma B(B^T\Sigma B + r^2)^{-1}$$

R_o : The solution of the matrix equation,

$$R_o = AR_oA^T + \sigma^2\;BB^T$$

we compute the asymptotic vector coefficients $\{b_i\}$, and the quantities in (61), as follows.

$$\underline{b}_i = [(I-\underline{\beta}\, B^T)A]^{-i}\, \underline{\beta} \quad ; \quad i=-m+1,\ldots,0$$

$$C = \sum_{i=0}^{m-1} [(I-\underline{\beta}\, B^T)A]^i \underline{\beta}\, B^T\, A^{-i}$$

$$S_u = R_o - A^m\, R_o (A^T)^m$$

$$S_{uz} = \sum_{i=0}^{m-1} [(I-\underline{\beta}\, B^T)A]^i\, \underline{\beta}\, B^T [R_o (A^T)^i - A^{m-i} R_o\, (A^T)^m]$$

$$S_z = r^2 \sum_{i=0}^{m-1} [(I-\underline{\beta}\, B^T)A]^i\, \underline{\beta}\, B^T [A^T(I-\underline{\beta}\, B^T)]^i +$$

$$+ \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} [(I-\underline{\beta}\, B^T)A]^i\, \underline{\beta}\, B^T [A^{u(j-i)} R_o\, (A^T)^{u(i-j)} -$$

$$- A^{m-i}\, R_o\, (A^T)^{m-j}]\, B\, \underline{\beta}^T\, [A^T(I-\underline{\beta}\, B^T)^T]^j$$

$$; \text{ where } u(i-j) = \begin{cases} i-j & ; \quad i \geq j \\ 0 & ; \quad \text{otherwise.} \end{cases}$$

We studied quantitative performance, for the following two special cases of the model in (57).

Model 1 First order autoregressive, with autoregressive parameter, $\alpha = 0.5$, and $\sigma^2 = r^2 = 1$.

Model 2 Third order autoregressive, with $a_1 = 0.6$, $a_2 = 0.07$, $a_3 = -0.06$ in (58), and $\sigma^2 = r^2 = 1$.

Tables 1, 2, and 3, and figures 1 and 2, exhibit the performance of the causal filtering operation in (59), for various values of the design parameters ε and m, when the nominal model is model 1. When the nominal model is instead model 2, the corresponding performance is exhibited by tables 4,5, and 6, and figure 3. Tables 2 and 5 correspond to independent per datum outliers, while tables 3 and 6 correspond to size-m independent batches of outliers. The above tables and figures speak for themselves. The causal filtering operation in (59) can combine close to optimal at the nominal model performance, together with excellent protection against outliers.

In addition, this operation is more appropriate for protection against independent batches of outliers. Similar results are drawn, when the order of the nominal model in (57) is some arbitrary number k.

We studied the performance of the nonrecursive causal operation in (41), when the nominal model is model 1. Our results are exhibited in tables 7, 8, and 9, and in figure 4, where various values of the design parameters ε and m are considered. Comparing the latter tables and figures, with tables 1, 2, and 3, and figures 1 and 2, we conclude that the nonrecursive causal operation in (41) induces higher asymptotic mean squared error at the nominal model, than the recursive operation in (59) does, while the former induces lower saturation point of the influence function. Considering this tradeoff, we claim that the recursive operation in (59) is more appropriate, for the autoregressive nominal model in (57).

We evaluated the performance of the smoothing operation in (42), for various orders of the autoregressive model in (57). Our results were similar to those exhibited in tables 1 to 6, and figures 1 to 4. The smoothing operation in (42) is thus as powerful as the filtering operation in (59).

9. Conclusions

We proposed and analyzed nonlinear filtering and smoothing operations, for effective resistance to outliers, and simultaneously good performance at the Gaussian nominal model. We note that a filtering operation, similar to our recursive such operation (in (59)), was earlier considered by Masreliez and Martin (1977). The latter authors assumed, however, that the process formed by the residuals is Gaussian, and used a covariance recursion to define their recursive filter. The above assumption effectively reduces the problem to the class of filters, that do not involve nested nonlinearities.

ε \ m	1	2	3	4	5	6
0.002	0.53167 0.66941	0.53284 0.56629	0.53333 0.54159	0.53346 0.53552	0.53350 0.53401	0.53351 0.53364
0.01	0.53488 0.67108	0.53963 0.57247	0.54136 0.54945	0.54183 0.54385	0.54195 0.54246	0.54198 0.54211
0.1	0.58157 0.70620	0.608293 0.63797	0.61640 0.62328	0.01851 0.62032	0.61904 0.61949	0.61917 0.61929
0.15	0.60983 0.72961	0.64401 0.67249	0.65401 0.66099	0.65659 0.65832	0.65723 0.65767	0.65740 0.65750
0.25	0.66941 0.78026	0.71376 0.73998	0.72608 0.73249	0.72921 0.73080	0.72999 0.73039	0.73019 0.73028
0.3	0.70079 0.80718	0.74848 0.77357	0.76146 0.76758	0.76474 0.76626	0.76556 0.76594	0.76576 0.76586
0.4	0.76727 0.86426	0.81887 0.84156	0.83243 0.83795	0.83582 0.83719	0.83667 0.83701	0.83688 0.83697

Table 1

Bounds on the asymptotic mean squared error, at the nominal model.
Model 1. Causal filtering operation in (59).
Asymptotic mean squared error induced by the optimal at the nominal model causal filter = 0.53112
Upper lines: lower bounds.

ε \ m	1	2	3	4	5	6
0.002	0.09928 0.14352	0.06814 0.07476	0.04932 0.05048	0.03788 0.03811	0.03056 0.03060	0.02556 0.02557
0.01	0.14699 0.20676	0.10040 0.10942	0.07274 0.07433	0.05597 0.05628	0.04522 0.04528	0.03786 0.03788
0.1	0.32204 0.40878	0.21602 0.22978	0.15715 0.15974	0.12180 0.12228	0.09898 0.09908	0.08326 0.08328
0.15	0.38225 0.47011	0.25595 0.27034	0.18674 0.18937	0.14516 0.14568	0.11824 0.11835	0.09962 0.09964
0.25	0.48129 0.56488	0.32349 0.33815	0.23761 0.24036	0.18576 0.18631	0.15194 0.15205	0.12840 0.12842
0.3	0.52466 0.60450	0.35423 0.36875	0.26119 0.26395	0.20477 0.20533	0.16783 0.16795	0.14203 0.14206
0.4	0.60417 0.67478	0.41329 0.42723	0.30739 0.31012	0.24246 0.24301	0.19955 0.19967	0.16937 0.16940

Table 2

Bounds on the breakdown point.
Model 1. Causal filtering operation in (59). Independent per datum outliers.
Upper lines: lower bounds.

ε \ m	1	2	3	4	5	6
0.002	0.09928 0.14352	0.13164 0.14394	0.14080 0.14394	0.14315 0.14394	0.14315 0.14394	0.14389 0.14394
0.01	0.14699 0.20676	0.19073 0.20686	0.20274 0.20682	0.20580 0.20682	0.20656 0.20682	0.20676 0.20682
0.1	0.32204 0.40878	0.38537 0.40676	0.40125 0.40653	0.40520 0.40651	0.40618 0.40651	0.40643 0.40651
0.15	0.38225 0.47011	0.44639 0.46759	0.46212 0.46733	0.46601 0.46731	0.46698 0.46731	0.46723 0.46731
0.25	0.48129 0.56488	0.54234 0.56195	0.55688 0.56166	0.56045 0.56164	0.56134 0.56164	0.56156 0.56164
0.3	0.52466 0.60450	0.58298 0.60152	0.56672 0.60124	0.60010 0.60121	0.60093 0.60121	0.60114 0.60121
0.4	0.60417 0.67478	0.65577 0.67193	0.66775 0.67166	0.67067 0.67164	0.67140 0.67164	0.67158 0.67164

Table 3

Bounds on the breakdown point.
Model 1. Causal filtering operation in (59). Independent size-m batches of outliers.
Upper lines: lower bounds.

ε \ m	1	2	3	4	5	6
0.002	0.55402 0.83214	0.57594 0.68407	0.59937 0.63361	0.61040 0.62154	0.61445 0.61764	0.61566 0.61658
0.01	0.57548 0.86214	0.62504 0.73994	0.66518 0.70200	0.68180 0.69383	0.68763 0.69109	0.68936 0.69035
0.1	0.62110 0.89436	0.69155 0.79589	0.72865 0.76040	0.74097 0.75110	0.74499 0.74788	0.74615 0.74698
0.15	0.65204 0.94013	0.72942 0.83110	0.74120 0.79568	0.77011 0.78320	0.77401 0.77516	0.77432 0.77501
0.25	0.69875 0.95182	0.73479 0.86264	0.76678 0.80203	0.78133 0.79312	0.79002 0.79202	0.79012 0.79136
0.3	0.73478 0.96067	0.73930 0.91011	0.78033 0.86481	0.79300 0.82414	0.80400 0.80923	0.80511 0.80547
0.4	0.73510 0.97033	0.74902 0.91437	0.79087 0.86690	0.81142 0.83571	0.82267 0.82610	0.82320 0.82359

Table 4

Bounds on the asymptotic mean squared error at the nominal model.
Model 2. Causal filtering operation in (59). Asymptotic mean squared error induced by the optimal at the nominal model causal filter = 0.54731.
Upper lines: lower bounds.

ε \ m	1	2	3	4	5	6
0.002	0.07594 0.13890	0.05802 0.07501	0.04513 0.04960	0.035395 0.035980	0.028765 0.029010	0.02411 0.02486
0.01	0.11029 0.20020	0.08225 0.11510	0.06334 0.08010	0.04958 0.05156	0.04030 0.0450	0.03380 0.03388
0.1	0.25689 0.39537	0.18640 0.22540	0.14353 0.15003	0.11313 0.11804	0.09248 0.09424	0.07790 0.07823
0.15	0.32899 0.47563	0.23552 0.27100	0.18083 0.20242	0.14286 0.14811	0.11705 0.11829	0.09878 0.09890
0.25	0.47094 0.60225	0.33123 0.39693	0.25350 0.26089	0.20121 0.20541	0.16563 0.16735	0.12747 0.12784
0.3	0.53838 0.65802	0.37811 0.42004	0.28952 0.29457	0.23047 0.23215	0.19020 0.19082	0.16150 0.16195
0.4	0.66166 0.75106	0.47002 0.52401	0.36191 0.39102	0.29019 0.30016	0.24090 0.24210	0.20548 0.20602

Table 5

Bounds on the breakdown point.
Model 2. Causal filtering operation in (59). Independent per datum outliers.
Upper lines: lower bounds.

ε \ m	1	2	3	4	5	6
0.002	0.07594 0.13890	0.11269 0.13995	0.12939 0.14500	0.13424 0.13952	0.13578 0.13595	0.13622 0.13682
0.01	0.11029 0.20020	0.15774 0.19500	0.17826 0.19851	0.18406 0.18820	0.18592 0.18683	0.18643 0.18682
0.1	0.25689 0.39537	0.33813 0.39220	0.37173 0.39104	0.38136 0.38804	0.38444 0.38740	0.38530 0.38607
0.15	0.32899 0.47563	0.41556 0.47215	0.45031 0.46903	0.46022 0.46630	0.46336 0.46502	0.46424 0.46482
0.25	0.47094 0.60225	0.55275 0.60112	0.58401 0.60039	0.59287 0.60004	0.59561 0.59970	0.59820 0.59918
0.3	0.53838 0.65802	0.61325 0.65720	0.64137 0.65695	0.64932 0.65530	0.65175 0.65398	0.65244 0.65307
0.4	0.66166 0.75106	0.71912 0.75083	0.74020 0.75010	0.74616 0.74970	0.74795 0.74912	0.74845 0.74887

Table 6

Bounds on the breakdown point.
Model 2. Causal filtering operation in (59). Independent size-m batches of outliers.
Upper lines: lower bounds.

m / ε	1	2	3	4	5	6
0.002	1.1434	1.1339	1.1334	1.1334	1.1334	1.134
0.01	1.1454	1.1360	1.1355	1.1355	1.1355	1.1355
0.1	1.1638	1.1553	1.1548	1.1548	1.1548	1.1548
0.15	1.1729	1.1648	1.1644	1.1644	1.1644	1.1644
0.25	1.1901	1.1830	1.1826	1.1826	1.1826	1.1826
0.3	1.1986	1.1918	1.1915	1.1915	1.1915	1.1915
0.4	1.2155	1.2096	1.2093	1.2092	1.2092	1.2092

Table 7

Asymptotoc mean squared error at the nominal model
Model 1. Causal filtering operation in (41).

m / ε	1	2	3	4	5	6
0.002	0.14395	0.07477	0.05048	0.03811	0.03060	0.02557
0.01	0.20683	0.10940	0.07433	0.05698	0.04528	0.03788
0.1	0.40652	0.22962	0.15963	0.12229	0.09909	0.08328
0.15	0.46731	0.27014	0.18937	0.14568	0.11835	0.09964
0.25	0.56164	0.33791	0.24036	0.18631	0.15206	0.12842
0.3	0.60122	0.36851	0.26394	0.20534	0.16795	0.14206
0.4	0.67165	0.42698	0.31011	0.24302	0.19967	0.16940

Table 8

Breakdown point.
Model 1. Causal filtering operation in (41). Independent per datum outliers.

m / ε	1	2	3	4	5	6
0.002	0.14395	0.14395	0.14395	0.14395	0.14395	0.14395
0. 01	0.20683	0.20683	0.20683	0.20683	0.20683	0.20683
0. 1	0.40652	0.40652	0.40652	0.40652	0.40652	0.40652
0. 15	0.46731	0.46731	0.46731	0.46731	0.46731	0.46731
0. 25	0.56164	0.56164	0.56164	0.56164	0.56164	0.56164
0. 3	0.60122	0.60122	0.60122	0.60122	0.60122	0.60122
0. 4	0.67165	0.67165	0.67165	0.67165	0.67165	0.67165

Table 9

Breakdown point.
Model 1. Causal filtering operation in (41).
Independent size-m batches of outliers.

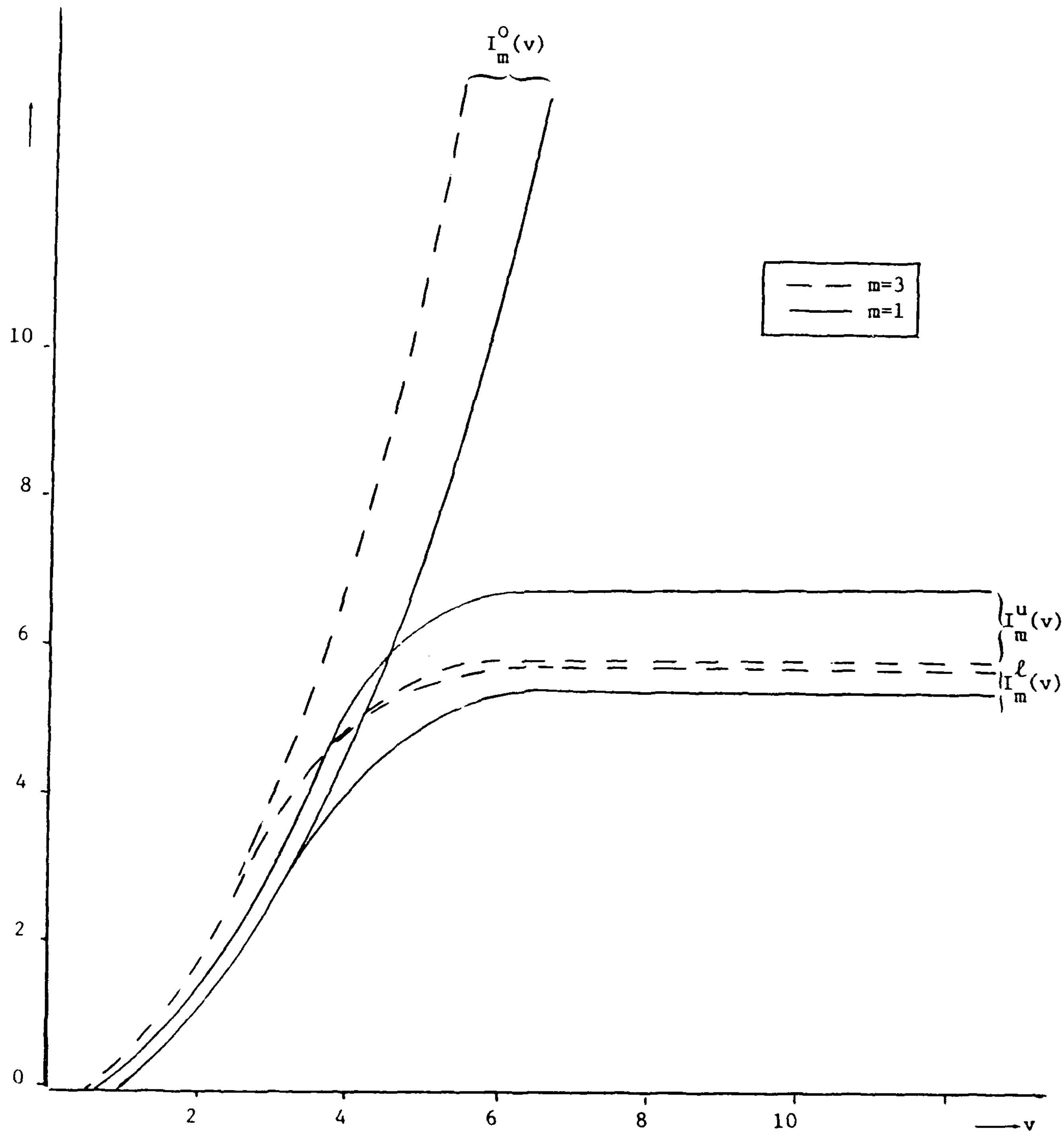


Figure 1

Bounds on the Influence Function
Model 1. Causal filtering operation in (59).
ϵ=0.002
$I_m^o(v)$: Influence function induced by the optimal at the nominal model filter.

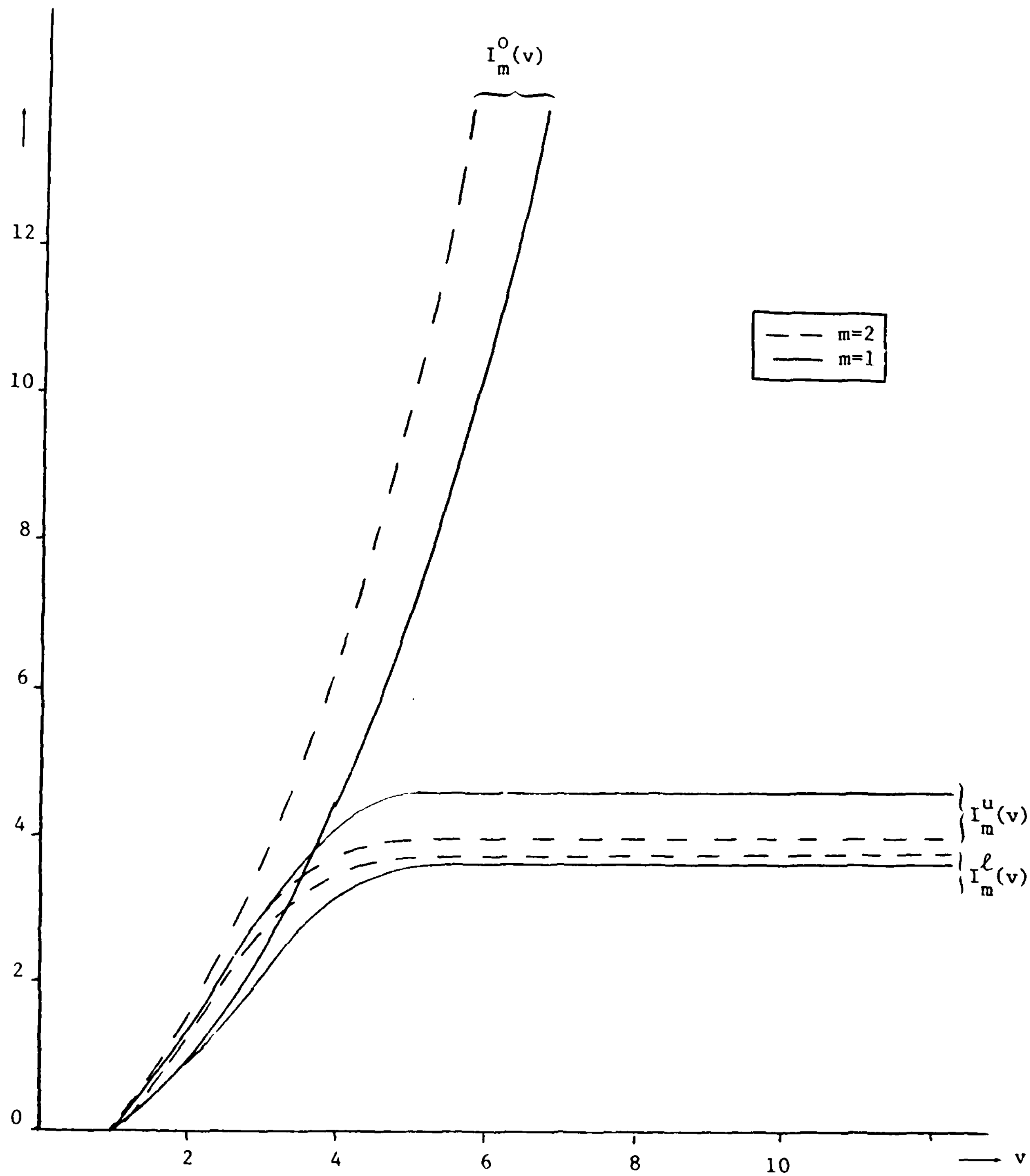


Figure 2

Bounds on the Influence Function
Model 1. Causal filtering operation in (59).
$\varepsilon=0.01$
$I_m^o(v)$: Influence function induced by the optimal at the nominal model filter.

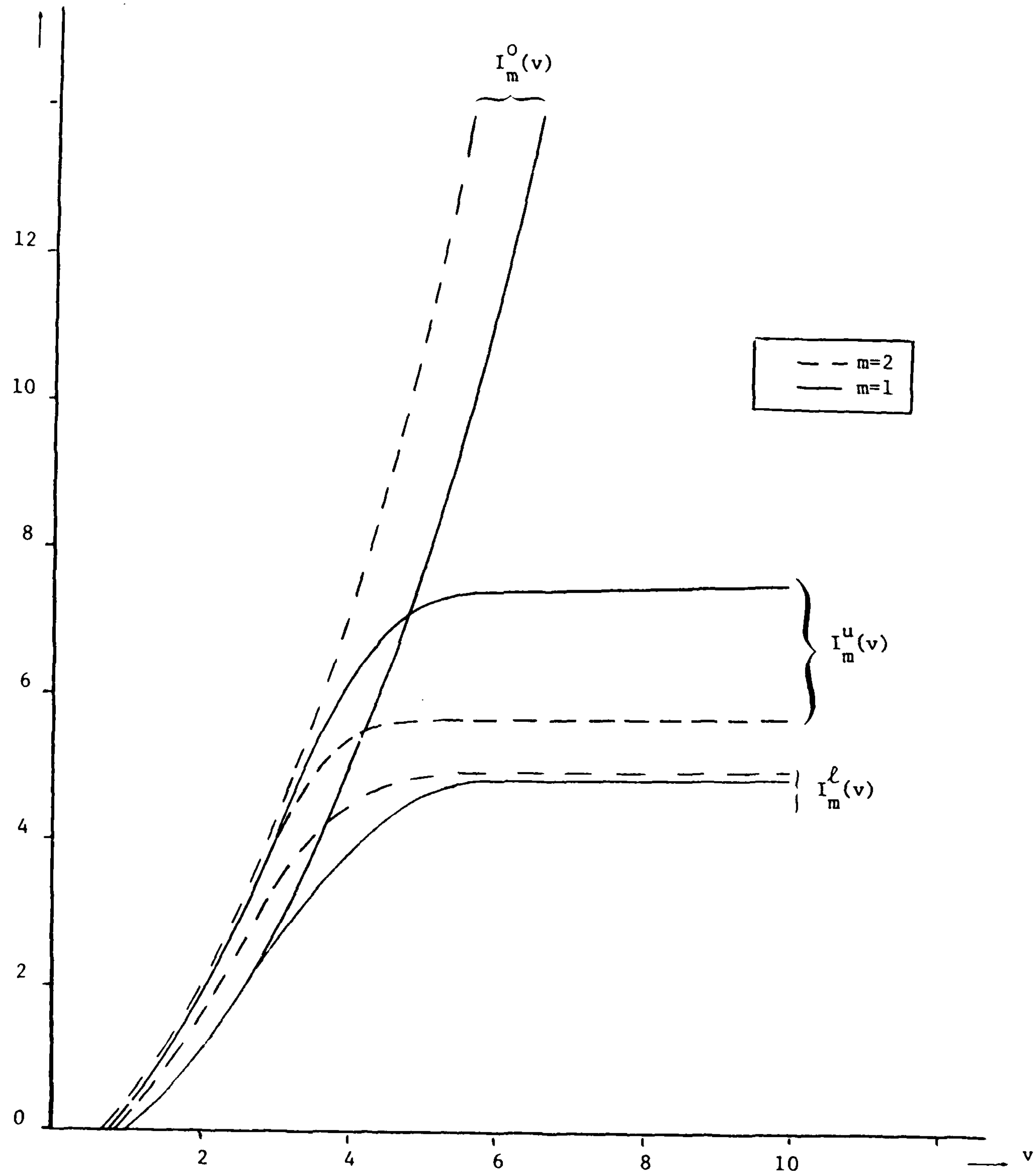


Figure 3

Bounds on the Influence Function
Model 2. Causal filtering operation in (59).
$\varepsilon=0.01$
$I_m^o(v)$: Influence function induced by the optimal at the nominal model filter.

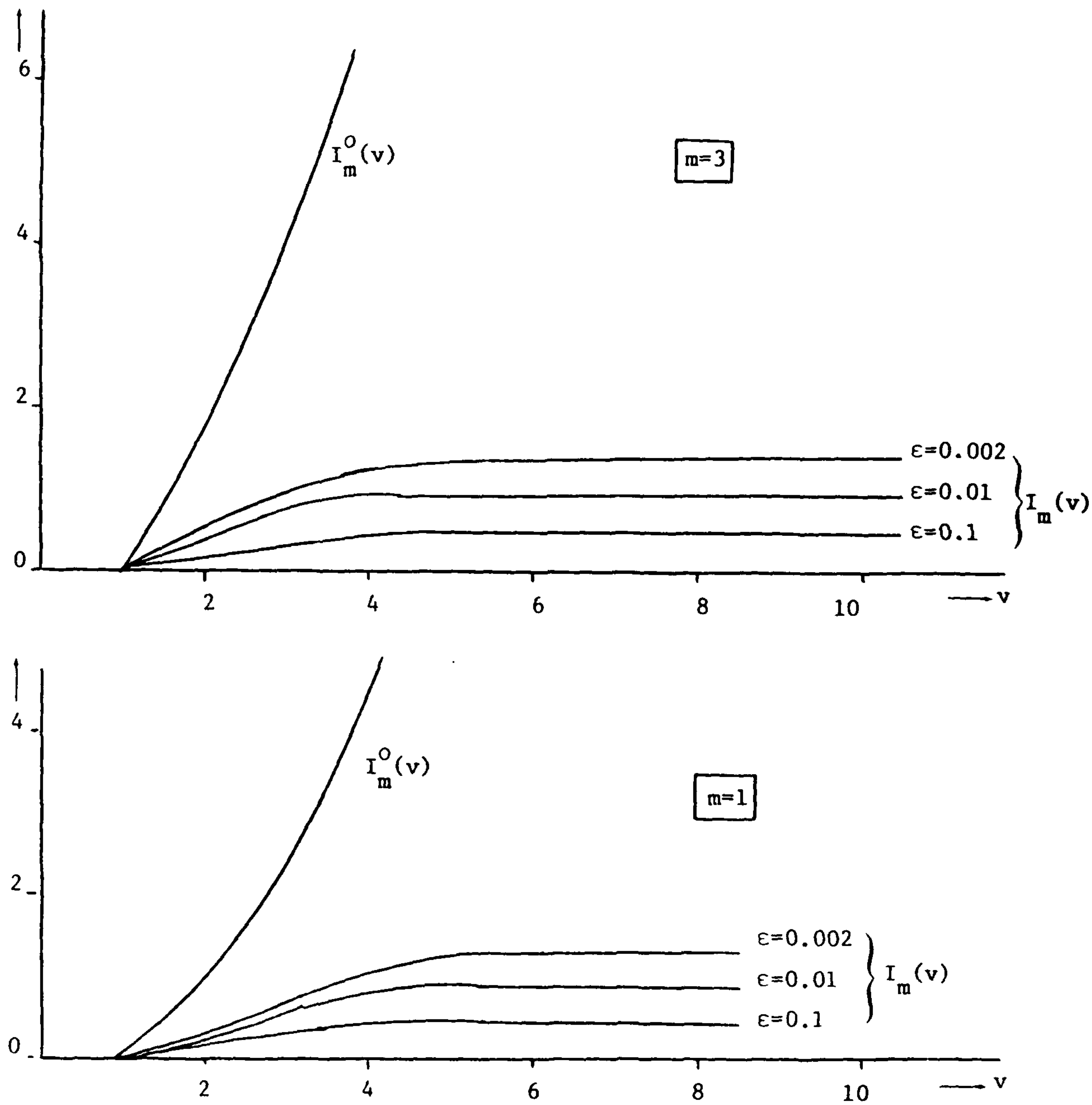


Figure 4

Influence Functions

Model 1. Causal filtering operation in (41).

$I_m^O(v)$: Influence function induced by the optimal at the nominal model filter.

Appendix A

Proof of Theorem 1

The class F^m and the functional $I(f^k)$ are both convex. Thus, for every δ : $0\leq\delta\leq 1$, we have,

$$\left((1-\delta)f_1^k + \delta\ f_2^k\ \varepsilon\right)F^k$$

$$I((1-\delta)f^k + \delta\ f_2^k\ ) \leq (1-\delta)\ \ I(f_1^k) + \delta\ I(f_2^k) = \inf_{f^m \varepsilon F^m}\ I(f^k) = \text{constant w.r.t.}\delta$$

Thus,

$$0= \frac{\partial^2\ I((1-\delta)f_1^k + \delta\ f_2^k)}{\partial\delta^2} = \int_{R^k} dy^k \left[\frac{\nabla_\ell f_1^k\ (y^k)}{f_1^k\ (y^k)} - \frac{\nabla_\ell f_2^k\ (y^k)}{f_2^k\ (y^k)}\right]^2 \frac{[f_1^k(y^k)f_2^k(y^k)]^2}{[(1-\delta)f_1^k(y^k)+\delta f_2^k(y^k)]^3}$$

from which we conclude the statement in the theorem, due to the continuity of the integrand.

Proof of Lemma 1

We wish to find some density function f_*^k in F^m that attains the infimum of the information measure $I(f^k)$ in (11), where the nominal density f_o^m in class F^m is the convolution of the Gaussian densities f_{os} and f_{oN}. Applying standard variational techniques, we conclude that if f_*^k satisfies the above infimum, then there exists some supset A^k of R^k, such that,

$$f_*^k(y^k) = (1-\varepsilon)\ f_o^k\ (y^k)\ ;\ \text{for}\ y^k\ \varepsilon\ A^k \tag{A.1}$$

For $y^k\ \varepsilon\ [R^k - A^k]$, the density $f_*^k\ (y^k)$ satisfies the following differential equation,

$$|\nabla_\ell f_*^k(y^k)| = \lambda f_*^k\ (y^k)\ ;\ \lambda > 0 \tag{A.2}$$

The general solution of the differential equation in (A.2) has the form,

$$f_*^k(y^k) = \exp\left\{ - \frac{\lambda |B^T y^k|}{B^T V_\ell^T} - C(y^k) \right\} ; y^k \epsilon\ [R^k - A^k] \tag{A.3}$$

; where B is any column vector in R^k that is not orthogonal to the vector V_ℓ^T, and where the scalar function $C(y^k)$ is such that, $\nabla_\ell C(y^k) = 0$. Imposing now continuity of the density function $f_*^k(y^k)$ and of its directional derivative $\nabla_\ell f_*^k(y^k)$, everywhere in R^k, we conclude,

$$A^k: \frac{|\nabla_\ell\ f_o^k(y^k)|}{f_o^k(y^k)} = |\ P_{ik}^T(\ell) y^k| \leq \lambda \tag{A.4}$$

For convenience, we now select $B = P_{ik}(\ell)$, in (A.3), where since M_{ik}^{-1} is positive definite, $P_{ik}(\ell)$ is not orthogonal to V_ℓ^T. From (A.3), (A.4), and (A.1), and requiring that $f_*^k(y^k)$ is continuous on the boundary $|P_{ik}^T(\ell) y^k| = \lambda$, we finally obrain the density in (14). The expressions (16) and (17) evolve from the requirement that $\int_{R^k} dy^k f^k{}_*(y^k) = 1$. The estimate in (15) is easily found to be the optimal at f_*^k mean squared estimate of the information datum x_ℓ.

Appendix B

Proof of Theorem 2

The operation in (30) has the general form,

$$\hat{x}_n = \sum_i a_i \hat{x}_i + g(y^m, \sum_i a_i \hat{x}_i, \{\hat{x}_j\}), \text{ where for some bounded } \lambda,\ g(x) = \begin{cases} x & ; |x| \le \lambda \\ \lambda \operatorname{sgn} x; & |x| > \lambda \end{cases},$$

and where $|\sum_i a_i| \le c$. Therefore, $|\hat{x}_n| \le \lambda[1 + |\sum_i a_i|] \le \lambda(c+1)$; $\forall n$. (B.1)

Let $E_{\mu_o}\{[X_n - \hat{x}_n]^2\}$ denote the mean squared error induced by the estimate $\hat{x}_n$, when the Gaussian nominal observation process is acting. Let $E_\mu\{[X_n - \hat{x}_n]^2\}$ be the same error, when some process in class F^m in (3) is acting instead. Let y^n and z^n denote sequences that are respectively generated by the processes μ_o and μ. Given some set A^n in R^n, and in conjuction with (B.1) and the Schwartz inequality, we have,

$$E_\mu\{[X_n - \hat{x}_n]^2 | z^n \varepsilon A^n\} = E_{\mu_o}\{X_n^2 | z^n \varepsilon A^n\} - 2E_\mu\{X_n \hat{x}_n | z^n \varepsilon A^n\} +$$

$$+ E_\mu\{[\hat{x}_n]^2 | z^n \varepsilon A^n\} \le$$

$$\le c + 2E^{1/2}\{X_n^2 | z^n \varepsilon A^n\} E^{1/2}\{[\hat{x}_n]^2 | z^n \varepsilon A^n\} + E_\mu\{[\hat{x}_n]^2 | z^n \varepsilon A^n\}$$

$$\le c + 2\lambda c^{1/2}(c+1) + \lambda^2(c+1)^2 = [c^{1/2} + \lambda(c+1)]^2 \stackrel{\Delta}{=} C \qquad \text{(B.2)}$$

Due to (B.2), and considering ergodic and stationary observation processes in conjuction with F^m, we obtain: Given $\eta > 0$, there exists n_o, such that,

$$\forall n > n_o; E_\mu\{[X_n - \hat{x}_n]^2\} \le (1 - \varepsilon + \eta) E\{[X_n - \hat{x}_n]^2 | [\#i : \gamma_m(z_{i+1}^{i+m}, y_{i+1}^{i+m}) > \varepsilon] \le n\varepsilon, y^n \varepsilon R^n\} + \varepsilon C$$

(B.3)

; where, for independent m-size outliers, there exists some $\varepsilon_o > 0$, such that,

$$E\{[X_n - \hat{x}_n]^2 | [\#i : \gamma_m(z_{i+1}^{i+m}, y_{i+1}^{i+m}) > \varepsilon] < n\varepsilon, y^n \varepsilon R^n\} \le$$

$$\leq E_{\mu_o}\{[X_n-\hat{x}_n]^2\} + \varepsilon C; \forall \varepsilon<\varepsilon_o, \forall n>n_o \qquad (B.4)$$

From (B.3) and (B.4), we conclude: Given $\eta=\frac{\varepsilon}{2}$, there exist n_o and $\varepsilon>0$, such that,

$$|E_\mu\{[X_n-\hat{x}_n]^2\}-E_{\mu_o}\{[X_n-\hat{x}_n]^2\}| \leq \varepsilon\ (2-\frac{\varepsilon}{2})C + \frac{\varepsilon}{2} E_{\mu_o}\{[X_n-\hat{x}_n]^2\} \leq$$

$$\leq \varepsilon\frac{5}{2} C \overset{\Delta}{=} \delta; \ \forall\ n>n_o,\ \forall\ \varepsilon<\varepsilon_o$$

Thus, given $\delta>0$, there exist, n_o and $\varepsilon: 0<\varepsilon<\min(\varepsilon_o, \frac{2}{5}\ \frac{\delta}{C})$, such that,

$$\Pi_{n,\rho_n}(\mu_o,\mu) < \varepsilon \text{ implies } |E_\mu\{[X_n-\hat{x}_n]^2\}-E_{\mu_o}\{[X_n-\hat{x}_n]^2\}|<\delta \quad ; \forall n>n_o$$

The proof of theorem is now complete.

<u>Proof of Theorem 3</u>

Both the operations in (32) and (39) have the form,

$$\hat{x}_{kA} = \sum_i a_i\ g\ (y_i^{i+m-1})$$

; where $|g(\cdot)|\leq\lambda$, and $(\sum_i a_i) \leq c$. Then, subject to any process, h, of mutually independent m-size batches of outliers, each occurring with probability ε, we have,

$$E_\mu\{[X_k-\hat{x}_{kA}]^2\} = E_{\mu_o}\{X_k^2\} -2\sum_i a_i\ E_\mu\{X_k g\ (Y_i^{i+m-1})\} +$$

$$+ \sum_i\sum_j a_i\ a_j\ E_\mu\{g(Y_i^{i+m-1})\ g\ (Y_j^{j+m-1})\}$$

$$= E_{\mu_o}\{X_k^2\} -2(1-\varepsilon) \sum_i a_i\ E_{\mu_o}\{X_k\ g(y_i^{i+m-1})\}-2\varepsilon \sum_i a_i\ E_h\{X_k g(Y_i^{i+m-1})\}$$

$$+ \sum_i\sum_j a_i\ a_j\ E_{\mu_o}\{g(Y_i^{i+m-1})g(Y_j^{j+m-1})\}$$

$$+ \sum_i\sum_j a_i\ a_j\ \left[E_\mu\{g(Y_i^{i+m-1})g(Y_j^{j+m-1})\} - E_{\mu_o}\{g(Y_i^{i+m-1})g(Y_j^{j+m-1})\}\right]$$

Or,

$$|E_\mu\{[X_k-\hat{x}_{kA}]^2\}-E_{\mu_o}\{[X_k-\hat{x}_{kA}]^2\}| \leq$$

$$\leq 2\varepsilon \sum_i a_i [E_{\mu_o}\{|X_k||g(Y_i^{i+m-1})|\} + E_h\{|X_k||g(Y_i^{i+m-1})|\}]$$

$$+ 4\varepsilon \sum_i \sum_j a_i \; a_j \; E_\mu\{|g(Y_i^{i+m-1})g(Y_j^{j+m-1})|\} \tag{B.5}$$

; where,

$$E_\mu\{|g(Y_i^{j+m-1})g(Y_j^{j+m-1})|\} \leq \lambda^2 < \infty \tag{B.6}$$

; and where applying the Schwartz inequality we have,

$$E_{\mu_o}\{|X_k||g(Y_i^{i+m-1})|\} \leq E_{\mu_o}^{1/2}\{X_k^2\}E_{\mu_o}^{1/2}\{g^2(Y_i^{i+m-1})\} \leq c^{1/2}\lambda < \infty$$

$$E_h\{|X_k||g(Y_i^{i+m-1})|\} \leq E_{\mu_o}^{1/2}\{X_k^2\}E_h^{1/2}\{g^2(Y_i^{i+m-1})\} \leq c^{1/2}\lambda < \infty \tag{B.7}$$

Applying inequalities (B.6) and (B.7) to (B.5), we obtain,

$$|E_\mu\{[X_k-\hat{x}_{kA}]^2\}-E_{\mu_o}\{[X_k-\hat{x}_{kA}]\}| \leq 4\varepsilon \; (\sum_i a_i) \; c^{1/2}\lambda \; + 4\varepsilon \; (\sum_i a_i)^2\lambda^2$$

$$= 4\varepsilon \; (\sum_i a_i)\lambda \; [c^{1/2} + \sum_i a_i] = \varepsilon C \tag{B.8}$$

; where $C \triangleq 4\lambda(\sum_i a_i) \; [c^{1/2} + \sum_i a_i] < \infty$.

Now, given $\delta > 0$, we select $\varepsilon = \delta \; C^{-1}$, to obtain,

$$\Pi_{n,\rho_n}(\mu_o,\mu) < \varepsilon = \delta \; C^{-1} \rightarrow |E_\mu\{[X_k-\hat{x}_{kA}]^2\} - E_{\mu_o}\{[X_k-\hat{x}_{kA}]^2\}| < \delta; \; \forall k$$

and the proof of the theorem is complete.

Proof of Theorem 4

We will prove the theorem in steps. Let r_p denote the pth autocorelation coefficient of either the sequence $\{Z_i = P_m^T Y_{i-m+1}^i\}$, or the sequence $\{E_i = P_n^T Y_{i-n}^{i+n}\}$, at the nominal Gaussian model. That is, either $r_p = E\{Z_i\, Z_{i+p} | f_o\}$, or $r_p = E\{E_i E_{i+p} | f_o\}$ Let then c be the constant in (26), and let us then denote,

$$\sigma^2 = r_0 \;,\; \gamma_p = \frac{r_p}{\sigma^2} \;,\; \sigma_p^2 = \sigma^2\, [1-\gamma_p^2] \tag{B.9}$$

Let L_p denote the pth autocorelation coefficient of either the sequence $\{g_{F,m}(Z_i)\}$, or the sequence $\{g_{S,n}(E_i)\}$, for $g_{F,m}(.)$ and $g_{S,n}(.)$ respectively as in (47) and (48), and at the nominal Gaussian model. That is, either

$L_p = E\{g_{F,m}(Z_i) g_{F,m}(Z_{i+p}) | f_o\}$, or $L_p = E\{g_{S,n}(E_i) g_{S,n}(E_{i+p}) | f_o\}$.

Let g(.) denote either $g_{F,m}(.)$ or $g_{S,n}(.)$, and let f(x) denote the zero mean, variance σ^2 Gaussian density function at the point x. Let λ denote the truncation threshold of the function g(.). Then, for $\phi(x)$ and $\Phi(x)$ being respectively the density and the distribution functions at the point x, of the zero mean unit variance Gaussian random variable, and for c as in (26), we directly obtain,

$$L_0 = \sigma^2\, \{2[1-c^2]\, \Phi\, (c) - 2c\phi(c) + 2c^2-1\} \tag{B.10}$$

$$L_p = 2\, \sigma_p \int_{-\infty}^{\infty} g(x) \left\{ \phi\left(\frac{\lambda+\gamma_p x}{\sigma_p}\right) + \frac{\lambda+\gamma_p x}{\sigma_p}\, \Phi\left(\frac{\lambda+\gamma_p x}{\sigma_p}\right) \right\} f(x)\, dx$$

$$-r_p\, [2\, \Phi(c)-1] \;;\; |p| \geq 1 \tag{B.11}$$

Denoting now by $\phi^{(k)}(x)$, the kth derivative of $\phi(x)$, at the point x, and applying the Taylor theorem, we have that,

Given n, there exists $t_n : 0 < t_n < 1$, such that,

$$\phi\left(\frac{\lambda+\gamma_p x}{\sigma_p}\right)+\frac{\lambda+\gamma_p x}{\sigma_p}\,\Phi\left(\frac{\lambda+\gamma_p x}{\sigma_p}\right)=$$

$$=\phi\left(\frac{\lambda}{\sigma_p}\right)+\frac{\lambda}{\sigma_p}\,\Phi\left(\frac{\lambda}{\sigma_p}\right)+\frac{\gamma_p x}{\sigma_p}\,\Phi\left(\frac{\lambda}{\sigma_p}\right)+$$

$$+\sum_{k=0}^{2(n-1)}\phi^{(k)}\left(\frac{\lambda}{\sigma_p}\right)\frac{\gamma_p^{k+2}\;x^{k+2}}{\sigma_p^{k+2}\;(k+2)!}\qquad+\phi^{(2n-1)}\left(\frac{\lambda}{\sigma_p}+t_n\frac{\gamma_p x}{\sigma_p}\right)\frac{\gamma_p^{2n+1}\;x^{2n+1}}{\sigma_p^{2n+1}\;(2n+1)!}$$

(B.12)

Substituting (B.12) in (B.11), and via some straightforward transformations, we obtain,

$$|p|\geq 1;\; L_p=2\sigma_p\int_{-\infty}^{\infty}g(x)\left\{\frac{\gamma_p x}{\sigma_p}\;\Phi\left(\frac{\lambda}{\sigma_p}\right)+\sum_{k=1}^{n-1}\phi^{(2k-1)}\left(\frac{\lambda}{\sigma_p}\right)\frac{\gamma_p^{2k+1}\;x^{2k+1}}{\sigma_p^{2k+1}\;(2k+1)!}+\right.$$

$$\left.+\phi^{(2n-1)}\left(\frac{\lambda}{\sigma_p}+t_n\frac{\gamma_p x}{\sigma_p}\right)\frac{\gamma_p{}^{2n+1}\;x^{2n+1}}{\sigma_p^{2n+1}(2n+1)!}\right\}f(x)dx\;-$$

$$-r_p[2\Phi(c)-1]=$$

$$=b_p(n)+2\frac{\gamma_p{}^{2n+1}}{\sigma_p^{2n}\;(2n+1)!}\int_{-\infty}^{\infty}x^{2n+1}g(x)f(x)\;\phi^{(2n-1)}\left(\frac{\lambda}{\sigma_p}+t_n\frac{\gamma_p x}{\sigma_p}\right)dx$$

(B.13)

;where, for c as in (26),

$$b_p(n)\triangleq r_p[2\Phi\left(\frac{c}{[1-\gamma_p^2]^{1/2}}\right)-1]\;[2\Phi(c)-1]$$

$$+\;2r_p\sum_{k=1}^{n-1}\left[\gamma_p^{-2}-1\right]^{-k}\phi^{(2k-1)}\left(\frac{c}{[1-\gamma_p^2]^{1/2}}\right)\left\{\frac{2\Phi(c)-1}{2^k\;k!}+\right.$$

$$\left.+\sum_{\ell=1}^{k}c^{2\ell-1}\;2^{k-\ell+1}\left[\frac{k!}{(2k+1)!\;(\ell-1)!}-2^{-2(k-\ell+1)}\frac{(\ell-1)!}{k!\;(2\ell-1)!}\right]\right\}\qquad(B.14)$$

$$\left| \int_{-\infty}^{\infty} x^{2n+1} g(x) f(x)\ \phi^{(2n-1)}\left(\frac{\lambda}{\sigma_p} + t_n \frac{\gamma_p x}{\sigma_p}\right) dx \right| <$$

$$\leq \left[\max_x \left| \phi^{(2n-1)}\left(\frac{\lambda}{\sigma_p} + t_n \frac{\gamma_p x}{\sigma_p}\right) \right| \right] \int_{-\infty}^{\infty} x^{2(n+1)} f(x) dx =$$

$$= C_n \frac{(2n+1)!}{2^n\ n!}\ \sigma^{2(n+1)} \qquad (B.15)$$

; where,

$$C_n = \max_x \left| \phi^{(2n-1)}(x) \right| < \frac{2^n\ n!}{\sqrt{2\pi\ e^{2n-1}}} \qquad (B.16)$$

From (B.13), (B.14), (B.15), (B.16), and (B.9), we conclude,

$$\left| L_p - b_p\ (n) \right| \leq r_p \left(\frac{2e}{\pi}\right)^{1/2} \left[\frac{\gamma_p^2}{e(1-\gamma_p^2)}\right]^n \triangleq B_p(n)\ ;\ |p| \geq 1 \qquad (B.17)$$

; where $b_p(n)$ is as in (B.14). We note that, if $\gamma_p^2 < \frac{e}{1+e}$, then given p, given $\zeta > 0$, there exists n_{op}, such that, $B_p(n) \leq r_p\zeta$; $\forall\, n > n_{op}$. Thus, if $\gamma_p^2 < \frac{e}{1+e}$; $\forall\, |p| \geq 1$, and if $\{r_p\}$ is such that, $r_{|p|} > r_{|p|+1} \searrow 0$, then given p, given $\zeta > 0$, there exists n_{op}, such that $B_p(n) \leq \zeta\, r_p$; $\forall\, n \geq n_{op}$, and $\{n_{op}\}$ is a decreasing sequence with increasing $|p|$. We thus conclude that, if $r_{|p|} > r_{|p|+1} \searrow 0$ and $\gamma_p^2 < \frac{e}{1+e}$; $\forall\, |p| \geq 1$, then, given $\zeta > 0$, there exist a decreasing with increasing $|p|$ sequence $\{n_{op}\}$, and a positive integer N, such that,

$$\begin{aligned} &b_p(n) - \zeta r_p \leq L_p \leq b_p(n) + \zeta r_p \quad ;\ \forall n \geq n_{op},\ \text{for given } p \\ &b_p(n) = b_p = r_p\ [2\Phi(c)-1]^2 \quad ;\ \forall\ p \geq N \end{aligned} \qquad (B.18)$$

Let us now denote by $C(\omega)$, either $D(\omega)$ in (45) or $A(\omega)$ in (46). Let us denote by $H(\omega)$, either $H_{F,m}(\omega)$ in (45) or $H_{S,n}(\omega)$ in (46). Let also $B(\omega)$ be the Fourier transform of the sequence $\{b_p(n_{op});|p| \leq N-1, b_p;|p| \geq N, L_0\}$ in (B.18) and (B.10). Let $e(f_o,\hat{x}_0)$ denote either the error $e(f_o,\hat{x}_{0F})$ in (45), or the error $e(f_o,\hat{x}_{0S})$ in (46). Then, due to (B.18) and (B.10), we directly obtain,

$$\left| e(f_o,\hat{x}_0) - e^{\ell}(f_o,\hat{x}_0) \right| \leq \zeta(2\pi)^{-4} \left| (2\pi) \int_{-\pi}^{\pi} ||C(\omega)||^2 f_s(\omega)\,d\omega - \left[\int_{-\pi}^{\pi} ||C(\omega)||^2 d\omega \right] \left[\int_{-\pi}^{\pi} f_s(\omega)\,d\omega \right] \right| \qquad \text{(B.19)}$$

; where,

$$e^{\ell}(f_o,\hat{x}_0) = (2\pi)^{-1} \left\{ \int_{-\pi}^{\pi} f_s(\omega)\,d\omega - 2[2\Phi(c)-1] \int_{-\pi}^{\pi} \mathrm{Re}(C(\omega))H(\omega))f_s(\omega)\,d\omega + \int_{-\pi}^{\pi} ||C(\omega)||^2 B(\omega)\,d\omega \right\} \qquad \text{(B.20)}$$

(B.20)

$$B(\omega) = \sigma^2 \{2[1-c^2]\,\Phi(c) - 2c\phi(c)+2c^2-1\} + \sum_{1\leq|p|\leq N-1} b_p(n_{op})e^{j\omega p} +$$

$$+ [2\Phi(c)-1]^2 \sum_{|p|\geq N} r_p e^{j\omega p} =$$

$$= [2\Phi(c)-1]^2 f_s(\omega) + \sigma^2 \{2\Phi(c)[3-c^2-2\Phi(c)] -2c\phi(c)$$

$$+2c^2-2\} + \sum_{1\leq|p|\leq N-1} e^{j\omega p}\{b_p(n_{op}) - [\Phi(c)-1]^2 \, r_p\} \tag{B.21}$$

The statement of the theorem follows directly from (B.19), (B.20), and (B.21).

Proof of Theorem 5

(i) The eigenvalues of the matrix A in (58) are the roots of the polynomial equation,

$$x^k - x^{k-1} a_1 - \ldots - x\, a_{k-1} - a_k = 0$$

Due to the assumed stationarity, their magnitudes are all less than one. Thus, the eigenvalues of the matrix A^{2m} have also magnitudes less than one.

Let us now consider the matrix C in (61) as a function of the design parameter m, and denote it then C_m. For m varying, we easily find that,

$$C_{m+1} = C_m + (I - C_m)\, A^m \, \underline{h}_{-m} \, B^T A^{-m} \tag{B.22}$$

; where,

$$\underline{h}_{-m} = \frac{\Sigma_{-m} B}{\sigma_{11}(-m) + r^2}$$

$$\Sigma_{-m} = \left\{ \sigma_{ij}(-m) \right\} = \tag{B.23}$$

$$= E\left\{\left[\underline{U}_{-m} - E\left\{\underline{U}_{-m} \mid y_{-\infty}^{-m-1}\right\}\right]\left[\underline{U}_{-m} - E\left\{\underline{U}_{-m} \mid y_{-\infty}^{-m-1}\right\}\right]^T\right\}$$

and where the expectations in (B.23) are taken at the nominal model. Also,

$$C_1 = \frac{\Sigma_0 \, B \, B^T}{\sigma_{11}(0) + r^2} \tag{B.24}$$

It is easily seen that the matrix C_1 in (B.24) has one eigenvalue between zero and one, and that the remaining eigenvalues equal zero. Due to (B.22), and by induction, the same assertion holds for C_m, $m > 1$. The magnitudes of the eigenvalues of the matrix Q are then also less than one, and the same assertion holds for the matrix P. Hence, the equations in (62) have solutions.

(ii) Let us consider the vector $\underline{Z}$ in (61), and let us define,

$$\begin{aligned} \underline{X} &\triangleq \underline{U}_0 - \underline{\hat{u}}_0 \\ \underline{\Omega} &\triangleq \underline{U}_{-m} - \underline{\hat{u}}_{-m} \\ \underline{Y} &\triangleq \underline{U}_0 - A^m \underline{U}_{-m} \end{aligned} \tag{B.25}$$

From (59) we then easily obtain,

$$\underline{X} = \underline{Y} + A^m \underline{\Omega} - \underline{g}_{F,m} (\underline{Z} + C A^m \underline{\Omega}) \tag{B.26}$$

Under the nominal model, the vectors $\underline{Y}$ and $\underline{Z}$ are jointly Gaussian, and independent of the vector $\underline{\Omega}$. Let $f_{\underline{X}}(\underline{x})$ and $f_{\underline{\Omega}}(\underline{\omega})$ denote the density functions of the vectors $\underline{X}$ and $\underline{\Omega}$ at the nominal model, and at respectively the vector points $\underline{x}$ and $\underline{\omega}$. Considering expectations at the nominal model, let us define,

$$S_{y/z} \triangleq E\left\{\left[\underline{Y} - E\left\{\underline{Y} \mid \underline{Z}\right\}\right]\left[\underline{Y} - E\left\{\underline{Y} \mid \underline{Z}\right\}\right]^T\right\}$$

$$R \triangleq S_{uz}^T \, S_z^{-1}$$

$$\Lambda\,(\underline{x},\,\underline{\omega}) = \int_{R^k} \frac{d\underline{z}\;\exp(-\frac{1}{2}\,\underline{z}^T\,S_z^{-1}\,\underline{z})}{(2\pi)^k\,|S_{y/z}|^{1/2}\,|S_z|^{1/2}}\;\exp\,(-\,\frac{1}{2}\,[\underline{x} + \underline{g}_{F,m}(\underline{z} + CA^m\underline{\omega})\; - A^m\underline{\omega}\; - R\underline{z}]^T.$$

$$S_{y/z}^{-1}\;[\underline{x} + \underline{g}_{F,m}(\underline{z} + CA^m\underline{\omega})\; - A^m\underline{\omega}\; - R\underline{z}]) \qquad (B.27)$$

; where S_{uz} and S_z are given by (61). In the sequel, we will use the following theorem, where for $\underline{x} = [x_1, x_2, \ldots, x_k]^T$, we define,

$$||\underline{x}|| \triangleq \max_i |x_i| \qquad (B.28)$$

Theorem B

Let $f(\underline{x},\underline{v}) : R^k \times R^\ell \rightarrow R^k$ be a measurable function, satisfying the following condition, for some positive real scalar function $h(\underline{v})$.

$$||f(\underline{x},\underline{v}) - f(\underline{x}',\underline{v})|| \leq ||\,\underline{x}-\underline{x}'|| \; h(\underline{v}) \;\; ; \forall\; \underline{x},\; \underline{x}'\epsilon R^k, \forall \underline{v}\epsilon R^\ell \qquad (B.29)$$

Let $\{\underline{X}_n,\; n \geq 0\}$ be a stochastic process in R^k, defined by the recursive equation,

$$\underline{X}_{n+1} = f(\underline{X}_n,\; \underline{V}_n) \quad , \; n \geq 0 \qquad (B.30)$$

; where $\{\underline{V}_n,\; n \geq 0\}$ is an i.i.d. process in R^ℓ, and where $\underline{V}_n$ is independent of X_n, for all $n \geq 0$. Let $\underline{V}_n$ be absolutely continuous, and let then $p(\underline{v})$ be its density function at $\underline{v}$. Let $p(\underline{v})$ be such that,

$$\int_{R^\ell} h(\underline{v})p(\underline{v})d\underline{v} = \zeta < 1 \qquad (B.31)$$

Then, the process $\{\underline{X}_n,\; n \geq 0\}$ is asymptotically stationary. Furthermore, the asymptotic distribution of $\underline{X}_n$, for $n \rightarrow \infty$ is the same for any initial distribution of X_0, and depends only on the function $f(\cdot,\cdot)$ and the density $p(\underline{v})$.

Proof

From (B.30) we conclude that $\{\underline{X}_n\}$ is a Markov process. Thus, to prove asymptotic stationarity, it suffices to show that, given any distribution for $\underline{X}_0$, the distribution of $\underline{X}_n$ converges weakly to a unique distribution in R^k, as $n \to \infty$.

Let $\mu_0(\underline{x})$ be an arbitrary density function, $\forall \underline{x} \varepsilon R^k$. Let then the sequence $\{\mu_n(\underline{x}),\ n \geq 0\}$ be defined as follows.

$$\mu_{n+1}(\underline{x}) = \int_{R^k} \Lambda(\underline{x},\underline{\omega})\ \mu_n(\underline{\omega})\ d\underline{\omega} \tag{B.32}$$

; where $\Lambda(\underline{x},\underline{\omega})$ denotes the conditional density function of $\underline{x}$, given $\underline{\omega}$, when $\underline{x} = f(\underline{\omega},\underline{v})$, and where $\underline{\Omega}$ is independent of $\underline{V}$, and $p(\underline{v})$ is the density function of $\underline{V}$ at $\underline{v} \varepsilon R^\ell$. Let us now define the sequence $\{\Lambda^{(n)}(\underline{x},\underline{\omega})\ ,\ n \geq 1\}$, as follows.

$$\begin{aligned} \Lambda^{(1)}(\underline{x},\underline{\omega}) &= \Lambda(\underline{x},\underline{\omega}) \\ \Lambda^{(n+1)}(\underline{x},\underline{\omega}) &= \int_{R^k} \Lambda^{(n)}(\underline{x},\underline{z})\ \Lambda^{(1)}(\underline{z},\underline{\omega})d\underline{z} \end{aligned} \tag{B.33}$$

Then, we can write,

$$\mu_n(\underline{x}) = \int_{R^k} \Lambda^{(n)}(\underline{x},\underline{\omega})\ \mu_0(\underline{\omega})\ d\underline{\omega} \tag{B.34}$$

To show weak convergence of the sequence $\{\mu_n(\underline{x})\}$, we need to show that there exists a density function $\mu(\underline{x})$; $\underline{x} \varepsilon R^k$, such that, for any continuous and bounded function, $g(\underline{x})$ in R^k, we have,

$$\int_{R^k} g(\underline{x})\ \mu_n(\underline{x})d\underline{x} \xrightarrow[n \to \infty]{} \int_{R^k} g(\underline{x})\ \mu(\underline{x})d\underline{x} \tag{B.35}$$

Let us define the sequence $\{g_n(\underline{x}),\ n \geq 0\}$, as follows.

$$g_0(\underline{x}) = g(\underline{x})$$
$$g_n(\underline{x}) = \int_{R^k} \Lambda(\underline{z},\underline{x})\ g_{n-1}(\underline{z})d\underline{z} \qquad (B.36)$$

Then,

$$\int_{R^k} g(\underline{x})\ \mu_n(\underline{x})d\underline{x} = \int_{R^k} g_n(\underline{x})\ \mu_0(\underline{x})d\underline{x} \qquad (B.37)$$

Let us define,

$$u_n \triangleq \sup_{\delta>0} \left(\delta^{-1} \sup_{||\underline{\omega}-\underline{z}||<\delta} |g_n(\underline{z}) - g_n(\underline{\omega})|\right) \qquad (B.38)$$

Without lack in generality, we will assume that the quantities $\{u_n\}$ are all finite (this is true if, for example, the functions $g_n(\underline{x})$ satisfy a Lipshitz condition). From (B.30) and (B.36), we obtain,

$$g_n(\underline{x}) = \int_{R^k} g_{n-1}(f(\underline{x},\underline{v}))\,p(\underline{v})d\underline{v} \qquad (B.39)$$

From (B.38) and (B.39), we conclude,

$$u_n \leq \int_{R^k} \left\{\sup_{\delta>0} \left(\delta^{-1} \sup_{||\underline{x}-\underline{\omega}||<\delta} |g_{n-1}(f(\underline{x},\underline{v})) - g_{n-1}(f(\underline{\omega},\underline{v}))|\right)\right\} p(\underline{v})d\underline{v}$$

$$\leq \int_{R^k} h(\underline{v})p(\underline{v}) \left\{\sup_{\delta>0} \left([\delta h(\underline{v})]^{-1} \sup_{||\underline{x}-\underline{\omega}||<\delta h(\underline{v})} |g_{n-1}(\underline{x}) - g_{n-1}(\underline{\omega})|\right)\right\} d\underline{v}$$

$$= u_{n-1} \int_{R^k} h(\underline{v})p(\underline{v})d\underline{v} = \zeta u_{n-1} < u_{n-1} \qquad (B.40)$$

From (B.40), we conclude that $u_n \to 0$, as $n \to \infty$, and that $g_n(\underline{x}) \to g(\underline{x})$ = constant on R^k, as $n \to \infty$. Thus,

$$\int_{R^k} g(x)\ \mu_n(x)dx = \int_{R^k} g_n(x)\mu_0(x)dx \xrightarrow[n\to\infty]{} \text{constant} \qquad (B.41)$$

Due to (B.29), the sequence $\{\mu_n(\underline{x})\}$ is tight. Thus, there exists a subsequence $\{\mu_{n_i}(\underline{x})\}$, and a density function $\mu(\underline{x})$ in R^k, such that, for every continuous and bounded function $g(\underline{x})$, we have,

$$\int_{R^k} \mu_{n_i}(\underline{x}) g(\underline{x}) d\underline{x} \xrightarrow[n_i \to \infty]{} \int \mu(\underline{x}) g(\underline{x}) d\underline{x} \tag{B.42}$$

From (B.41) and (B.42), it immediately follows that,

$$\int_{R^k} g(\underline{x}) \mu_n(\underline{x}) d\underline{x} \xrightarrow[n \to \infty]{} \int_{R^k} g(\underline{x}) \mu(\underline{x}) d\underline{x}$$

and the proof of the theorem is now complete.

Let us apply now theorem B, for expression (B.26). In this case,

$$\underline{V} = [\underline{Y}, \underline{Z}]^T \quad , \ \ell = 2k$$

$$f(\underline{\Omega}, \underline{V}) = \underline{Y} + A^m \underline{\Omega} - g_{F,m}(\underline{Z} + C\, A^m \underline{\Omega})$$

; where, for $|\mu_{max}(B)|$ denoting the absolutely largest eigenvalue of B, we have,

$$\| f(\underline{X}, \underline{V}) - f(\underline{\Omega}, \underline{V}) \| \leq \| A^m (I-C)(\underline{X} - \underline{\Omega}) \| \leq$$

$$\leq |\mu_{max} (A^m[I-C])| \; \| \underline{X} - \underline{\Omega} \| \tag{B.43}$$

Since $|\mu_{max}(A^m[I-C])| < 1$, the conditions in theorem B are clearly satisfied, with $h(\underline{v}) \triangleq |\mu_{max}(A^m[I-C])|$. Thus, the densities $f_{\underline{X}}(\underline{x})$ and $f_{\underline{\Omega}}(\underline{\omega})$ are asymptotically identical, where due to (B.26) and (B.27), we have,

$$f_{\underline{X}}(\underline{x}) = \int_{R^k} \Lambda(\underline{x}, \underline{\omega})\, f_{\underline{\Omega}}(\underline{\omega})\, d\underline{\omega} \tag{B.44}$$

From (B.44) we then obtain,

$$S_{oF} = E\{\underline{X}\, \underline{X}^T\} = \int_{R^k} M(\underline{\omega})\, f_{\underline{\Omega}}(\underline{\omega})\, d\underline{\omega} \tag{B.45}$$

; where,

$$M(\underline{\omega}) \triangleq \int_{R^k} \underline{x}\, \underline{x}^T \Lambda(\underline{x}, \underline{\omega})\, d\underline{x} = \{M_{ij}(\underline{\omega})\} \tag{B.46}$$

The function $M(\underline{\omega})$ in (B.46) is analytic, and possesses Taylor exnantion. In particular, applying the Taylor theorem we write,

$$M_{ij}(\underline{\omega}) = M_{ij}(\underline{0}) + [\nabla^T M_{ij}(\underline{\omega})|\underline{\omega} = \underline{0}]\underline{\omega} + 2^{-1}\,\underline{\omega}^T[\nabla^2 M_{ij}(\underline{\omega})|\underline{\omega} = \underline{\mu}]\underline{\omega} \;;\; \text{for some } \underline{\mu}:$$

$$\underline{0} \leq \underline{\mu} \leq \underline{\omega} \quad \text{(B.47)}$$

; where we can easily find that,

$$M(\underline{0}) = P$$

$$\text{(B.48)}$$

$$\underline{\omega}^T (A^T)^m A^m\, \underline{\omega} - \underline{\omega}^T\, (A^T)^m\, Q_{ij}\, A^m\, \underline{\omega} \leq \underline{\omega}^T\, [\nabla^2\, M_{ij}(\underline{\omega})\,|\, \underline{\omega} = \underline{\mu}]\, \underline{\omega} \leq \underline{\omega}^T\, (A^T)^m\, A^m\, \underline{\omega} \;;\; \forall\, \underline{\omega}$$

The expressions in (B.48), in conjuction with the equality $\int \underline{\omega} f_{\underline{\Omega}}(\underline{\omega})d\underline{\omega} = 0$, and expressions (B.45) and (B.47), give the inequality in (63).

(iii) Via substitution, we find,

$$\left|tr\,(S^u_{oF} - S^{\ell}_{oF})\right| \leq \left|tr\,(P + \sum_{i=1}^{\infty} A^{mi}\, P\, (A^T)^{mi} - P)\right| =$$

$$= \left|tr\,(A\,A^T)^m\,(I-[AA^T]^m)^{-1}\,P\right| \leq \frac{|\mu_{max}(A)|^{2m}}{1-|\mu_{max}(A)|^{2m}}\, tr\, P \qquad \text{(B.49)}$$

; where the inequality in (B.49) evolves from the fact that the magnitudes of the eigenvalues of the matrix P are all less than one.

Appendix C

Proof of Theorem 6

Let $e^{\ell}_{N_F,\{n_{pF}\},F}(f_o,\delta,\hat{x}_0)$ and $e^{\ell}_{N_S,\{n_{pS}\},S}(f_o,\delta,\hat{x}_0)$ denote respectively the mean squared errors $e(f_o,\delta,\hat{x}_{0F})$ and $e(f_o,\delta,\hat{x}_{0S})$, when the correlation coefficients $r_{g_{F,m}}(k)$ and $r_{g_{S,n}}(k)$ are substituted by the bounds $b_{kF}(n)$ and $b_{kS}(n)$, in (52). Then, we easily find,

$$e^{\ell}_{N_F,\{n_{pF}\},F}(f_o,\delta,\hat{x}_0) = E\{X_0^2|f_o\} + F_F(\delta)$$

$$e^{\ell}_{N_S,\{n_{pS}\},S}(f_o,\delta,\hat{x}_0) = E\{X_0^2|f_o\} + F_S(\delta) \tag{C.1}$$

$$e^{\ell}_{N_A,\{n_{pA}\},A}(f_o,0,\hat{x}_0) + E\{X_0^2|f_o\} + F_A(0) = e^{\ell}_{N_A,\{n_{pA}\},A}(f_o,\hat{x}_0);\ A=F,S$$

; where $e^{\ell}_{N_A,\{n_{pA}\},A}(f_o,\hat{x}_0)$; A=F,S are the error bounds in theorem 4.

It can be easily seen from (70), (71), and (73), that $e(f_o,\delta,\hat{x}_{0F})$, $e(f_o,\delta,\hat{x}_{0S})$, $F_F(\delta)$, and $F_S(\delta)$ are all monotonically increasing with increasing δ, and are all continuous and differentiable as functions of δ. Also,

$$e(f_o,0,\hat{x}_{0F}) < E\{X_o^2|f_o\},\ e(f_o,0,\hat{x}_{0S}) < E\{X_0^2|f_o\}, \text{ and,}$$

$$e(f_o,1,\hat{x}_{0F}) = E\{X_0^2|f_o\} + \lambda_m^2 D^2(0)$$

$$e(f_o,1,\hat{x}_{0S}) = E\{X_0^2|f_o\} + \mu_n^2 A^2(0) \tag{C.2}$$

Thus, the breakdown points $\delta^*_{F,m}$ and $\delta^*_{S,n}$ exist, are unique, and are strictly between zero and one. In addition, if can be easily found that the functions $G_F(\delta) \triangleq e(f_o,\delta,\hat{x}_{0F}) - E\{X_0^2|f_o\} - F_F(\delta)$, and $G_S(\delta) \triangleq e(f_o,\delta,\hat{x}_{0S}) - E\{X_0^2|f_o\} - F_S(\delta)$ are such that,

$$G_F(1) = G_S(1) = 0$$

$$|G_F(\delta)| < |G_F(0)| = |e(f_o,\hat{x}_{0F}) - e^{\ell}_{N_F,\{n_{pF}\},F}(f_o,\hat{x}_0)| \qquad (C.3)$$

$$|G_S(\delta)| \leq |G_S(0)| = |e(f_o,\hat{x}_{0S}) - e^{\ell}_{N_S,\{n_{pS}\},S}(f_o,\hat{x}_0)|$$

Due to (C.3), and via theorem 4, given $\zeta_1 > 0$, we can find integers N_F and N_S, and sets $\{n_{pF};1 \leq |p| \leq N_F-1\}$, $\{n_{pS};1 \leq |p| \leq N_S-1\}$, such that, $|G_F(\delta)| < \zeta_1;\forall\delta$, $|G_S(\delta)| \leq \zeta_1;\forall\delta$. Then, for ζ_1 small enough, we can also have, $F_F(0) < 0$ and $F_S(0) < 0$, while at the same time we have, $F_F(1)=\lambda_m^2 D^2(0) > 0$ and $F_S(1)=\mu_n^2 A^2(0) > 0$. Thus, the functions $F_F(\delta)$ and $F_S(\delta)$ will then have unique zeros, $\delta^o_{F,m}$ and $\delta^o_{S,n}$. Due to the continuity of all the functions, $e(f_o,\delta,\hat{x}_{0F})$ $e(f_o,\delta,\hat{x}_{0S})$, $F_F(\delta)$, and $F_S(\delta)$, with respect to δ, given $\zeta > 0$, we can find $\zeta_1 > 0$, such that $|G_F(\delta)| < \zeta_1;\forall\delta$ and $|G_S(\delta)| < \zeta_1;\forall\delta$, gives:

$$|\delta^*_{F,m}-\delta^o_{F,m}| < \zeta \text{ and } |\delta^*_{S,n}-\delta^o_{S,n}| < \zeta.$$

<u>Proof of Lemma 2</u>

Let us define, $\underline{Y}$, $\underline{X}$, and $\underline{\Omega}$ as in (B.25), and let us also define,

$$\underline{Z}_1 \triangleq \sum_{i=-m+1}^{0} \underline{b}_i \, B^T [\, \underline{U} \;\; -A^{m+i} \; \underline{U}_{-m}] \qquad (C.4)$$

$$\underline{Z}_2 \triangleq \sum_{i=-m+1}^{0} \underline{b}_i \, W_i$$

Then, for C as in (61), the optimal at the nominal model filtering operation can be written as follows.

$$\underline{X} = \underline{Y} + A^m \underline{\Omega} - [\underline{Z}_1 + \underline{Z}_2 + C A^m \underline{\Omega}] \qquad (C.5)$$

; where $\underline{\Omega}$ is independent of $\underline{Y}$, $\underline{Z}_1$, and $\underline{Z}_2$, where $\underline{Y}$ and $\underline{Z}_1$ are independent of $\underline{Z}_2$, and where the vectors $\underline{Y}$ and $\underline{Z}_1$ are zero mean and jointly Gaussian. In addition, denoting by $G(M,\Sigma)$ the mean M and covariance matrix Σ Gaussian distribution, the vector $\underline{Z}_2$ has the following distribution, in the persence of the m-size block outlier model.

$$(1-\delta)\ G\ (0,r^2N) + \delta\ G\ (v\underline{\mu},0) \quad ; \quad 0 < \delta < 1 \qquad (C.6)$$

Due to (C.6) and the fact that the distributions of $\underline{X}$ and $\underline{\Omega}$ are asymptotically identical, we easily find from (C.5),

$$E\{\underline{X}\} = -\delta v\ [I - (I - C)\ A^m]^{-1}\ \underline{\mu} \qquad (C.7)$$

From (C.5) and (C.7), we also find,

$$\begin{aligned} E\{\underline{X}\ \underline{X}^T\} = {} & (I-C)\ A^m\ E\{\underline{X}\ \underline{X}^T\}\ (A^T)^m\ (I-C)^T + \\ & + (1-\delta)\ (E\{(\underline{Y}-\underline{Z}_1)\ (\underline{Y}-\underline{Z}_1)^T\} + r^2\ N) + \\ & + \delta(E\{(\underline{Y}-\underline{Z}_1)\ (\underline{Y}-\underline{Z}_1)^T\} + v^2\ \underline{\mu}\ \underline{\mu}^T) + \\ & + \delta^2 v^z((I-C)\ A^m\ [I-(I-C)\ A^m]^{-1}\ \underline{\mu}\ \underline{\mu}^T + \\ & + \underline{\mu}\ \underline{\mu}^T\ [I-\ (A^T)^m\ (I-C)^T]^{-1}\ (A^T)^m\ (I-C)^T) \qquad (C.8) \end{aligned}$$

Solving (C.8), we finally find, for $E_o\{\underline{X}\ \underline{X}^T\} = E\{\underline{X}\ \underline{X}^T\}$ at $\delta=0$,

$$\begin{aligned} I_m^o\ (v) &= \lim_{\delta\to 0} \frac{E\{\underline{X}\ \underline{X}^T\} - E_o\{\underline{X}\ \underline{X}^T\}}{\delta} = \\ &= \sum_{i=0}^{\infty} [(I-C)\ A^m]^i\ [v^2\ \underline{\mu}\ \underline{\mu}^T - r^2N]\ [(A^T)^m(I-C)^T]^i \qquad (C.9) \end{aligned}$$

; where the infinite sum in (C.9) converges, since the eigenvalues of $(I-C)A^m$ have magnitudes that are strictly less than one.

Proof of Theorem 7

(i) Let $\underline{X}$, $\underline{Y}$, and $\underline{\Omega}$ be defined as in (B.25), appendix B, and considering expectations at the nominal model, let us define,

$$S_{y/z_1} \triangleq E\{[\underline{Y} - E\{\underline{Y}|\underline{Z}_1\}]\ [\underline{Y} - E\{\underline{Y}|\underline{Z}_1\}]^T\}$$

$$R_1 \triangleq S_{uz}^T\ S_{z_1}^{-1}$$

$$\Lambda_v(\underline{x},\underline{\omega}) \triangleq (2\pi)^{-k}\ |S_{y/z_1}|^{-1/2}\ |S_{z_1}|^{-1/2} \int_{R^k} d\underline{z}\ \exp(-2^{-1}\underline{z}^T S_{z_1}^{-1}\underline{z}). \tag{C.9}$$

$$.\exp(-2^{-1}[\underline{x} + \underline{g}_{F,m}(\underline{z}+v\underline{\mu}+CA^m\underline{\omega})-A^m\underline{\omega}-R_1\underline{z}]^T\ S_{y/z_1}^{-1}[\underline{x} + \underline{g}_{F,m}(\underline{z}+v\underline{\mu}+CA^m\underline{\omega})-A^m\underline{\omega}-R_1\underline{z}])$$

Given δ, and the m-size block outlier model at level v, the vectors $\underline{Y}$, $\underline{Z}$, and $\underline{Z}_1$ are still independent of the vector $\underline{\Omega}$. Then, for $f_{\underline{X}}(\underline{x})$ and $f_{\underline{\Omega}}(\underline{\omega})$ respectively denoting the density functions of $\underline{X}$ and $\underline{\Omega}$, for $\Lambda(\underline{x},\underline{\omega})$ defined as in (B.27), appendix B, and for,

$$\Lambda_{\delta,v}(\underline{x},\underline{\omega}) \triangleq (1-\delta)\Lambda(\underline{x},\underline{\omega}) + \delta\Lambda_v(\underline{x},\underline{\omega}) \tag{C.10}$$

we can write,

$$f_{\underline{X}}(\underline{x}) = \int_{R^k} \Lambda_{\delta,v}(\underline{x},\underline{\omega})\ f_{\underline{\Omega}}(\underline{\omega})\ d\underline{\omega} \tag{C.11}$$

It can be easily found that theorem B, in the proof of theorem 5, appendix B, applies on (C.11) with $h(\underline{v}) \triangleq |\mu_{max}(A^m[I-C])| < 1$. Thus, the densities $f_{\underline{X}}(\underline{x})$ and $f_{\underline{\Omega}}(\underline{\omega})$ in (C.11) are asymptotically identical. From (C.11), we then obtain,

$$E\{\underline{X}\ \underline{X}^T\} = \int_{R^k} [(1-\delta)\ M(\underline{\omega}) + \delta M_v(\underline{\omega})]\ f_{\underline{\Omega}}(\underline{\omega})\ d\underline{\omega} \tag{C.12}$$

; where $M(\underline{\omega})$ is as in (B.46), appendix B, and is bounded as in (B.48), and where,

$$M_v(\underline{\omega}) \triangleq \int_{R^k} \underline{x}\ \underline{x}^T\ \Lambda_v(\underline{x},\underline{\omega})\ d\underline{x} \tag{C.13}$$

Applying the Taylor theorem on the analytic matrix function $M_v(\underline{\omega})$, and similarly to (B.47), for the matrix $M(\underline{\omega})$, we find the following expression, where $A \leq B$ means that each diagonal element of the matrix A is bounded from above by the corresponding diagonal element of the matrix B.

$$P_v + A^m \underline{\omega}\,\underline{\omega}^T (A^T)^m + [\nabla M_v(\underline{\omega})|_{\underline{\omega}=\underline{0}}]\,\underline{\omega} \geq M_v(\underline{\omega}) \geq P_v + A^m \underline{\omega}\,\underline{\omega}^T (A^T)^m +$$
$$+ [\nabla M_v(\underline{\omega})|_{\underline{\omega}=\underline{0}}]\underline{\omega} - R_v(\underline{\omega}) \qquad \text{(C.14)}$$

; where $R_v(\underline{\omega}) = \{\underline{\omega}^T (A^T)^m Q_{ij,v} A^m \underline{\omega}\ ;\ i,j = 1,\ldots,k\}$. Let us now define $\delta^o_{\underline{X}}(\underline{x})$ and $\delta^o_{\Omega}(\underline{\omega})$ as follows.

$$\delta^o_{\underline{X}}(\underline{x}) = \int_{R^k} \Lambda(\underline{x},\underline{\omega})\ \delta^o_{\Omega}(\underline{\omega})\ d\underline{\omega} \qquad \text{(C.15)}$$

; where from the proof of theorem 5, we have that $\delta^o_{\underline{X}}(\underline{x})$ and $\delta^o_{\Omega}(\underline{\omega})$ are asymptotically identical. Then, considering also the densities in (C.11), we easily find,

$$I_m(v) = \int_{R^k} M(\underline{\omega}) \left[\lim_{\delta\to 0} \delta^{-1}(\delta_{\Omega}(\underline{\omega}) - \delta^o_{\Omega}(\underline{\omega})) \right] d\underline{\omega} + \int_{R^k} [M_v(\underline{\omega}) - M(\underline{\omega})]\delta^o_{\Omega}(\underline{\omega})\ d\underline{\omega} \qquad \text{(C.16)}$$

; where from theorem 5 we have,

$$\int_{R^k} \underline{\omega}\ \delta^o_{\Omega}(\underline{\omega})\ d\underline{\omega} = 0 \quad , \quad S^{\ell}_{OF} \leq \int_{R^k} M(\underline{\omega})\ \delta^o_{\Omega}(\underline{\omega})\ d\underline{\omega} \leq S^u_{OF}$$

and where,

$$\int_{R^k} \underline{\omega}\ \underline{\omega}^T [\lim_{\delta\to 0} \delta^{-1}(\ \delta_{\Omega}(\underline{\omega}) - \delta^o_{\Omega}(\underline{\omega}))]\ d\underline{\omega} = I_m(v)$$

From the above, and using (C.14) on (C.16), we finally find the result in (80).

(ii) The results follow easily from the bounds in (B.47), appendix B, and the bounds in (C.14), by setting $v \to \infty$, and by substituting $1-\delta$ by $(1-\delta)^m$, and δ by $1-(1-\delta)^m$. In particular, we then find,

$$f_2(\delta) \triangleq tr(S^{\ell}_m(\delta) - S_o) \leq tr(E\{\underline{X}\ \underline{X}^T\} - E\{\underline{U}_0\ \underline{U}_0^T\}) \leq tr(S^u_m(\delta) - S_o) \triangleq f_1(\delta) \qquad \text{(C.17)}$$

; where the functions $f_1(\delta)$ and $f_2(\delta)$ are trivially found to have unique, strictly positive, and strictly less than one roots, δ^u_m and δ^{ℓ}_m. The inequalities in (82) follow then trivially from (C.17).

References

Boente, G., Fraiman, R., and Yohai, V., (1982),"Qualitative robustness for general stochastic processes", Dept. of Statistics, Univ. of Washington Seattle, Tech. Report No. 26, Oct.

Bustos, O., Fraiman, R., and Yohai, V., (1984), "Asymptotic behavior of the estimates based on residual autocovariances for ARMA models", Robust and nonlinear time series analysis, in Lecture Notes in Statistics, 26-49, Springer-Verlag, N.Y.

Cox, D., (1978), "Metrics on Stochastic processes and qualitative robustness", Dept. of statistics, Univ. of Washington Seattle, Tech. Report No. 23.

Hampel, F. R., (1971), "A general qualitative definition of robustness", Annals Math. Stat., 42, 1887-1895.

Hampel, F. R., (1974), "The influence curve and its role in robust estimation", J. Amer. Statist. Assoc., 69, 383-394.

Huber, P.J., (1981), Robust Statistics, Wiley.

Kalman, R. E., (1960), "A new approach to linear filtering and prediction problems", Journal of Basic Engineering (ASME), Vol. 82D, pp. 35-45.

Kalman, R. E., (1963), "New methods in Weiner filtering", Chapter 9, Proc. of the First Symposium on Engineering Applications of Random Function Theory and Probability, John Whiley and Sons. Inc., New York.

Kolmogorov, (1941), "Stationary sequences in Hilbert space", Bull. Math. Univ., Moscow, 2, 40.

Martin, R. D., and Yohai, V. J., (1984), "Influence curves for time series," Univ. of Washington, Dept. of Statistics, Technical Report No. 51, June.

Martin, R. D., and Zeh, J. E., (1977), "Determining the character of time series outliers", Proceedings of the American Statistical Assoc.

Masreliez, C. J., and Martin, R. D., (1977), "Robust bayesian estimation for the linear model and robustifying the Kalman filter", IEEE Trans. on Aut. Control, AC-22, 361-371.

Papantoni-Kazakos, P., and Gray, R. M., (1979), "Robustness of estimators on stationary observations", Ann. Probability, 7, 989-1002.

Papantoni-Kazakos, P., (1981), "Sliding block encoders that are Rho-Bar continuous functions of their input", IEEE Trans. Inf. Theory, IT-27, 372-376.

Papantoni-Kazakos, P., (1983), "Qualitative robustness in time series", EECS Dept., Univ. of Connecticut, Tech. Report UCT/DEECS/TR-83-15, Nov.

Papantoni-Kazakos, P., (1984a), "A game theoretic approach to robust filtering", Information and Control, 60, Nos. 1-3, 168-191.

Papantoni-Kazakos, P., (1984b), "Some aspects of qualitative robustness in time series", Robust and Nonlinear Time Series Analysis, in Lecture Notes in Statistics, 218-230, Springer-Verlag, N.Y.

Whittle, P., (1953), "The analysis of multiple stationary time series", J. Roy. Statist. Soc., Ser. B., 15, 125-139.

Wiener, N., (1949), Extrapolation, interpolation and smoothing of stationary time series, Cambridge MIT Press.

END
DTIC
8-86